AD_____

Award Number: W81XWH-13-1-0360

TITLE: Targeting One-Carbon Metabolism in Breast Cancer

PRINCIPAL INVESTIGATOR: Elizaveta Freinkman

CONTRACTING ORGANIZATION: Whitehead Institute for Biomedical Research
Cambridge, MA 02142

REPORT DATE: April 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2014	Final	15 September 2013 – 31 March 2014

4. TITLE AND SUBTITLE
Targeting One-Carbon Metabolism in Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-1-0360

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Elizaveta Freinkman

E-Mail: freinkm@wi.mit.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Whitehead Institute for Biomedical Research
Cambridge, MA 02142

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The altered metabolism of cancer cells is important for their viability, growth and proliferation, and targeting such metabolic alterations is a validated strategy for ablating tumor cells while sparing normal tissue. However, little is known about the metabolic requirements underlying cancer cell aggressiveness – a phenotype that includes increased drug resistance, invasiveness, stem-like properties, and metastatic potential, and is often characterized by an epithelial-to-mesenchymal transition (EMT) in cellular identity. Triple-negative and metastatic breast cancers are particularly aggressive, lack effective therapies, and therefore carry a poor survival prognosis. To identify targetable metabolic requirements of these cancers, we developed a complete metabolomic profiling platform including optimized methods for sample preparation, liquid chromatography/mass spectrometry (LC/MS), and untargeted data analysis. Furthermore, using a cell-culture model, we identified the metabolic enzyme dihydropyrimidine dehydrogenase (DPYD) as specifically required for the EMT. These results showcase the capability of our LC/MS-based metabolomics platform to measure small molecules in the context of cancer metabolism; in addition, we identify DPYD as the first metabolic enzyme specifically required for the EMT, a process associated with the acquisition of aggressive characteristics in breast cancer and other carcinomas. This work led to a manuscript (Appendix 2) now under revision for publication in *Cell*.

15. SUBJECT TERMS
breast cancer, carcinoma, aggressiveness, metastasis, epithelial-to-mesenchymal transition, metabolism, pyrimidine, stem-like

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	62	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

The altered metabolism of cancer cells is important for their viability, growth and proliferation, and targeting such metabolic alterations is a validated strategy for ablating tumor cells while sparing normal tissue. However, little is known about the metabolic requirements underlying cancer cell aggressiveness – a phenotype that includes increased drug resistance, invasiveness, stem-like properties, and metastatic potential, and is often characterized by an epithelial-to-mesenchymal transition (EMT) in cellular identity. Triple-negative and metastatic breast cancers are particularly aggressive, lack effective therapies, and therefore carry a poor survival prognosis. By using a cell-culture model of the EMT, we sought to understand the critical metabolic requirements that may reflect targetable liabilities of these deadly cancers.

Keywords

breast cancer, carcinoma, aggressiveness, metastasis, epithelial-to-mesenchymal transition, metabolism, pyrimidine, stem-like

Overall Project Summary

Task 1. Mapping the metabolome of aggressive breast cancer *in vitro* and in patient samples

To perform this untargeted metabolomic analysis, we first needed to develop appropriate data collection and analysis strategies. Accordingly, we have developed two liquid chromatography methods with overlapping coverage of many classes of polar metabolites, and we have validated these methods and obtained retention time (RT) data using approximately 150 chemical standards (see Appendix 1). For analysis of amino acids and central carbon metabolites, we found that the optimal method utilized polymeric hydrophilic interaction liquid chromatography (ZIC-pHILIC analytical column, 2.1x150 mm, 5 μm particle size, Merck) with a 30 min. gradient from 80% to 20% acetonitrile against an aqueous buffer containing 20 mM ammonium carbonate and 0.1% ammonium hydroxide. For analysis of sugars, nucleobases, and organophosphates, we found that the optimal method utilized a Luna amino column (2.0x150 mm, 3 μm particle size, Phenomenex) with a 20 min. gradient from 90% to 10% acetonitrile against an aqueous buffer containing 5 mM ammonium acetate and 0.2% ammonium hydroxide. We also adapted a third LC/MS method, originally described by Bird et al. (ref. 1), for the analysis of lipids. We routinely operate our instruments in polarity switching mode in order to maximize the number of detected metabolites.

In developing our sample preparation strategy, we compared six metabolite extraction protocols for cultured cells (relevant to Task 1a) and four protocols for tissues, using mouse liver samples (relevant to Task 1b). These protocols were gathered from the literature and from consultation with other metabolomics research groups; details of the protocols are summarized in Tables 1 and 2. After comparing the metabolite signal intensities for these protocols (example data shown in Figure 1), we adopted the ice-cold 80% methanol protocols (C-1 and T-1) for experiments in which only polar metabolites are to be analyzed because it is rapid and provides good yield of a wide variety of metabolites; for experiments in which lipids as well as polar metabolites are to be analyzed, the more laborious chloroform-methanol extraction protocols (C-3 and T-3) will be used, and the two phases will be analyzed separately.

In addition to these method development efforts, we also obtained all necessary approvals for the analysis of human breast tumor and matched normal tissue, as planned in Task 1b.

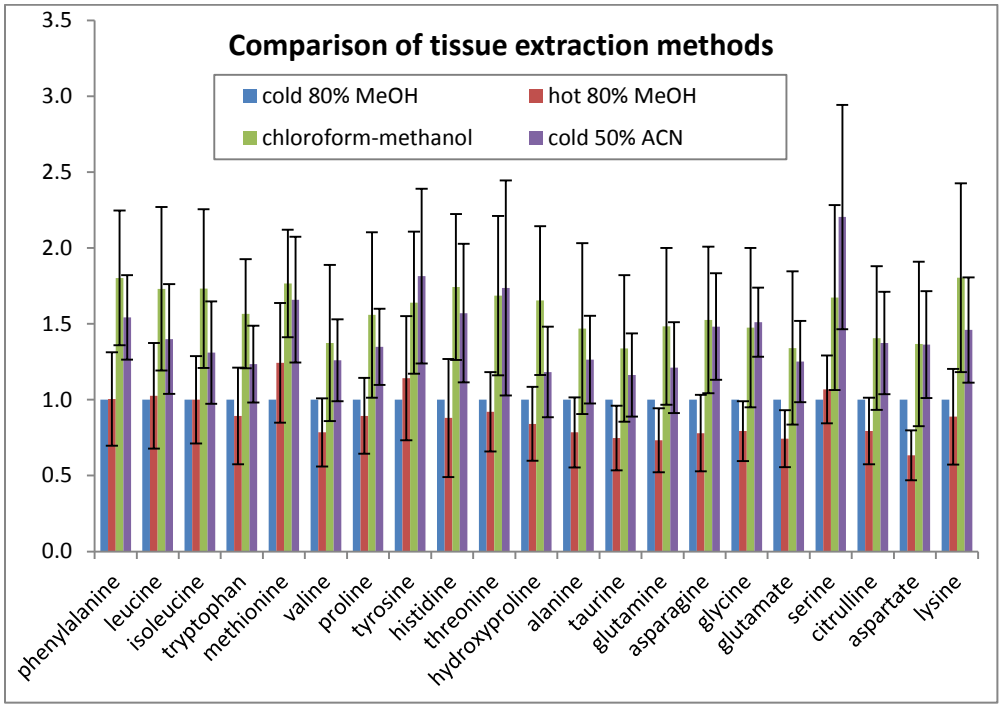
Table 1. Summary of cellular extraction protocols tested for Task 1a. Abbreviations: MeOH, methanol; ACN, acetonitrile; iPrOH, isopropanol. See references 2-5.

	C-1	C-2	C-3	C-4	C-5	C-6
wash	2x1.5ml 0.9% NaCl, 0°C	1x1ml H_2O, 0°C; liquid N_2 quench	1x1.5ml 0.9% NaCl, 0°C	1x1.5ml 0.9% NaCl, 0°C	1x1.5ml 0.9% NaCl, 0°C	2x1.5ml 0.9% NaCl, 0°C
extraction	1ml 80% MeOH, 0°C	200ul 40:40:20 ACN/MeOH/ 200mM NaCl, 10 mM Tris-HCl, pH 9.2	600ul MeOH, 300ul H2O, 400ul $CHCl_3$, -20°C	250ul 0.9% NaCl, 250ul MeOH, 500ul $CHCl_3$, -20°C	1ml 5:3:2 MeOH/ACN/ H_2O, -20°C	350ul 80% MeOH, 400ul $CHCl_3$, -20°C
post-extraction	dry, resuspend in H_2O	analyze immediately	separate layers, dry, resuspend in H_2O (polar) or 65:30:5 ACN/iPrOH/ H_2O (nonpolar)	separate layers, dry, resuspend in H_2O (polar) or 65:30:5 ACN/iPrOH/ H_2O (nonpolar)	analyze immediately	separate layers, dry, resuspend in H_2O (polar) or 65:30:5 ACN/iPrOH/ H_2O (nonpolar)

Table 2. Summary of tissue extraction protocols tested for Task 1b. All extraction solvents were used at 1 ml per 10-30 mg frozen tissue. Abbreviations are as in Table 1. See reference 6.

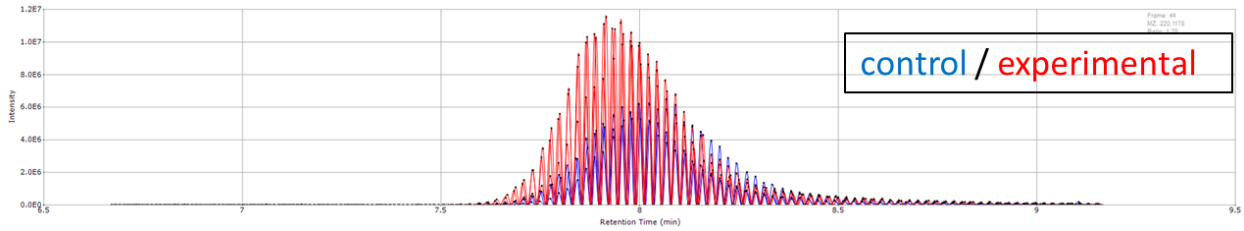
	T-1	T-2	T-3	T-4
extraction	80% MeOH, 0°C	80% MeOH, 70°C	1:1:1 $CHCl_3$:H_2O:MeOH	50% ACN, 0°C
post-extraction	bead homogenization, separation of phases (T-3 only), drying and resuspension			

Figure 1. Comparison of amino acid levels detected in 6 mouse liver samples extracted by four distinct protocols (levels were normalized to the cold 80% methanol protocol, T-1).



An additional prerequisite for Task 1 was to establish appropriate data analysis methods for comparing the metabolite content of human mammary epithelial cells that have or have not undergone an EMT (Task 1a), or of breast tumor vs. normal tissue (Task 1b), in an untargeted manner. To do this, we used the commercially available software packages Progenesis CoMet and SIEVE, which identify LC/MS peaks that differ significantly between sample groups. To ensure that the detected peaks are significantly different, we established the following quality-control cutoffs: (1) $p<0.05$ for differences between sample groups; (2) peak width >0.1 min; and (3) m/z error limit of 5 ppm for any metabolite identifications associated with a given peak. Figure 2 shows an example in which SIEVE identified an LC/MS peak that significantly differed in abundance between a control and an experimental group of cultured cell samples (used for data analysis optimization purposes only). The fold change in abundance and the metabolite identification – pantothenic acid – were confirmed by manual re-analysis of the original data and by comparison with an authentic standard of this metabolite. These data indicate that our data analysis methods can successfully discover significant metabolomic differences as proposed in Task 1.

Figure 2. Example of untargeted metabolomic data analysis by the software package SIEVE. Each curve represents an individual extracted ion chromatogram for m/z=220.1178 in a control (blue) or experimental (red) cell sample (both control and experimental groups were analyzed in biological triplicate). The m/z and retention times of this peak matched those of pantothenic acid, consistent with the metabolite identification provided by SIEVE.



Task 2. Identifying metabolic drivers of aggressive breast cancer

We used a bioinformatic approach (Task 2a) to identify candidate metabolic enzymes whose inhibition may prevent the EMT. These metabolic enzymes were labeled as the "mesenchymal metabolic signature" (MMS) genes because they were highly expressed in high-grade, mesenchymal-like carcinomas relative to low-grade, epithelial-like counterparts of the same tissue of origin (see Figures 1 and 2 of Appendix 2). We then performed a FACS-based shRNA screen to identify MMS genes whose knockdown prevented HMLE cells from undergoing EMT (Appendix 2, Figure 3A). Among the MMS genes, the gene dihydropyrimidine dehydrogenase (DPYD) was a top hit (Appendix 2, Figures 3B-D).

Because no chemical inhibitors of DPYD were available (Task 2b), we assessed the effects of DPYD knockdown on EMT *in vitro* (Task 2c). Indeed, knockdown of DPYD robustly inhibited EMT as judged by expression of EMT-specific cell-surface markers and transcription factors and by the mammosphere formation assay, a measure of mesenchymal character (Appendix 2, Figure 4). These effects were not due to non-specific cellular toxicity, as cellular proliferation was not affected (Appendix 2, Figure 4D). Furthermore, we showed that the catalytic activity of DPYD is required for EMT (Appendix 2, Figure 6A-F) and that the products of DPYD enzymatic activity, the dihydropyrimidines, increase in abundance during EMT (Appendix 2, Figure 5) and can substitute for DPYD activity when added to cell culture media (Appendix 2, Figure 6G). These results demonstrated the utility of our LC/MS platform for

measuring small molecules in the context of cancer metabolism; furthermore, with these results, we identified DPYD as the first metabolic enzyme specifically required for EMT, a process associated with the acquisition of aggressive traits in breast cancer and other carcinomas. This work led to a manuscript (Appendix 2) now under revision for publication in *Cell*.

Key Research Accomplishments

- Optimized sample preparation, LC/MS, and data analysis methods for use in untargeted metabolomic analysis of epithelial-to-mesenchymal transition and of aggressive breast tumors relative to normal breast tissue
- Identified dihydropyrimidine dehydrogenase (DPYD) as upregulated in high-grade carcinomas and essential for epithelial-to-mesenchymal transition (EMT) in human mammary epithelial cells
- Demonstrated an essential role for DPYD enzymatic activity and its products (dihydropyrimidines) in EMT

Conclusion

Our results suggest that the metabolic enzyme dihydropyrimidine dehydrogenase (DPYD) may be a useful diagnostic marker and/or a drug target in aggressive carcinomas, such as triple-negative and metastatic breast cancers. Measuring the expression levels of this enzyme in tumors may help predict their metastatic potential, while inhibition of this enzyme may limit or prevent metastasis. To explore these possibilities, current efforts include ablating DPYD expression in an animal model of metastatic breast cancer in order to determine the effects on metastasis.

Publications, Abstracts, and Presentations

a. Manuscripts

Shaul Y.D., Freinkman E., Comb W.C., Cantor J.R., Tam W.L., Thiru P., Kim D., Pacold M.E., Chen W.W., Bierie B., Possemato R., Weinberg R.A., Yaffe M.B., Sabatini D.M. DPYD is a key component of a metabolic program required for the epithelial-mesenchymal transition. Under revision, *Cell*.

b. Presentations

Freinkman E. "Metabolomics at the Whitehead Institute Small Molecule Analysis Center." Poster presentation, Whitehead Institute annual retreat, Sept. 2013.

Shaul Y.D. "DPYD is a key component of a metabolic program required for the epithelial-mesenchymal transition." Poster presentation, Keystone Symposium on Tumor Metabolism (X6), Mar. 2014.

Inventions, Patents, and Licenses

None to report

Reportable Outcomes

We identified the metabolic enzyme DPYD as essential for the epithelial-to-mesenchymal transition (EMT), a process associated with the acquisition of tumor drug resistance and metastasis. This finding suggests that DPYD expression level in a tumor may be a useful

diagnostic marker of metastasis risk, and that pharmaceutical inhibition of DPYD may limit tumor aggressiveness and metastasis.

Other Achievements

The research supported by this award and the expertise gained by Dr. Freinkman as a BCRP postdoctoral fellow has led to the establishment of a new research facility, the Metabolite Profiling Core Facility, at the Whitehead Institute for Biomedical Research (WI). This facility, which will be directed by Dr. Freinkman effective April 1, 2014, will provide LC/MS-based metabolite profiling on a collaborative basis. This experimental capability, which was previously unavailable at WI, is already expanding the scope of biomedical research being performed here, including important fundamental studies relevant to cancer and infectious disease.

Opportunities for Training and Professional Development

I have learned a tremendous amount during the fellowship period. Scientific mentoring and development occurred through biweekly one-on-one meetings with my mentor; weekly lab meetings (where I presented my research in February, August and October of 2013) and floor meetings with other MIT laboratories working in the cancer field; monthly subgroup meetings; and other regular, relevant seminars such as those of the MIT Biology Department, the Koch Institute for Integrative Cancer Research, and the Broad Institute Metabolism Initiative. I also received extensive training in the use and maintenance of metabolomics instrumentation (LC/MS and GC/MS) as well as metabolomics data analysis with software including Thermo XCalibur, LCQuan, XCMS, and Progenesis CoMet. This occurred through on-site training as well as off-site and Web-based seminars such as "Lipidomics and LipidSearch Software," "Small Molecule Structural Elucidation and Unknown Characterization," and "Thermo Annual Mass Spec User Meeting" (a scientific meeting organized by Thermo Fisher, the manufacturer of our LC/MS instruments). I also participated extensively in the writing and editing of a manuscript (Shaul, Y.D., et al.) currently under revision at *Cell*.

References

1. Bird et al., *Anal. Chem.* 2011, 83, 6648-6657.
2. Fendt et al., *Nature Commun*. 2013, 4, 2236.
3. Yang et al., *Cell* 2014, 156, 317-331.
4. Homan et al., *J. Am. Chem. Soc.* 2011, 133, 5178-5181.
5. Maddocks et al., *Nature* 2013, 493, 542-546.
6. Dietmair et al., *Anal. Biochem.* 2010, 404, 155-164.

pHILIC method

	RT (min)	polarity
2-aminoadipic acid	15.63	+
2'-deoxyadenosine	6.48	+
2'-deoxycytidine	8.99	both
2'-deoxyguanosine	9.77	both
2'-deoxyinosine	8.03	both
2'-deoxyuridine	6.65	-
2-hydroxyglutarate	16.03	-
2-ketobutyrate	5.87	-
3,5-diiodotyrosine	14.31	both
4-hydroxy-2-oxoglutarate	16.93	-
4-hydroxyisoleucine	11.71	+
5-hydroxyindole-3-acetate	12.70	both
5-hydroxytryptophan	13.79	both
acetyl-CoA	13.41	+
adenine	7.90	+
adenosine	7.28	+
alanine	14.59	+
arginine	21.83	+
asparagine	15.14	both
aspartate	15.80	both
citrulline	15.83	both
cystathionine	16.95	both
cysteine	8.74	+
cystine	16.62	+
cytidine	10.72	+
cytosine	9.84	+
dAMP	13.69	both
dGMP	16.86	both
dihydrofolate	16.96	+
dihydroorotate	10.78	-
dUMP	15.08	both
folic acid	17.93	both
folinic acid	19.09	both
glucuronic acid	16.52	-
glucuronic acid γ-lactone	8.78	-
glutamate	15.59	both
glutamine	14.99	+
glutathione, oxidized (GSSG)	18.59	both
glutathione, reduced (GSH)	15.20	both
glycine	15.59	+
GMP	18.04	both
guanine	11.16	both
histidine	14.40	+
homocysteine	16.06	+
homoserine	14.64	+
hydroxyproline	14.55	+
isocitrate	19.48	-
isoleucine	10.59	+
itaconic acid	15.80	-

Luna method

	RT (min)	polarity
1-methyladenosine	5.48	+
2'-deoxyadenosine	4.66	+
2'-deoxycytidine	6.05	both
2'-deoxyguanosine	7.65	both
2'-deoxyinosine	8.78	-
2'-deoxyuridine	4.43	-
2-hydroxyglutarate	10.73	-
2'-O-methyladenosine	3.36	+
3,5-diiodotyrosine	11.98	both
3-hydroxyphenylacetic acid	8.63	-
5-hydroxyindole-3-acetate	11.38	-
AMP	11.63	both
ATP	14.32	both
biotin	9.12	both
blasticidin	9.28	both
CDP	13.04	both
cis -aconitic acid	11.94	-
citrate	11.98	-
CMP	11.60	both
creatinine	5.15	both
crotonoyl-CoA	12.55	+
CTP	14.31	-
cytidine	6.73	both
cytosine	5.85	both
dADP	12.42	both
dAMP	11.53	both
dATP	14.26	both
dCDP	12.89	both
dCMP	11.51	both
dCTP	13.54	both
dGDP	13.94	both
dGMP	12.43	both
dGTP	15.12	both
diaminopimelic acid	10.48	both
dihydrofolate	13.00	both
dihydroorotate	8.79	-
dUMP	11.56	both
dUTP	14.35	-
folic acid	12.70	-
folinic acid	11.60	-
fructose	7.22	-
fructose 1-phosphate	10.96	-
fructose 6-phosphate	11.39	-
fructose-1,6-bisphosphate	13.48	-
fumarate	10.70	-
galactose	7.36	-
galactose 1-phosphate	11.37	-
GDP	14.10	both
geranyl pyrophosphate	11.84	-

ketoisoleucine	4.75	-
ketoleucine	4.69	-
kynurenine	9.44	+
lactate	8.86	-
leucine	9.98	+
lysine	21.12	+
malate	16.77	-
melatonin	4.41	+
methionine	10.86	+
methylglyoxal	14.18	+
mevalonate	7.87	-
N,N-dimethylglycine	11.34	+
N-acetyl-5-hydroxytryptamine	5.04	+
N-acetylglutamate	14.92	+
N-acetylphenylalanine	4.86	both
N-acetylserine	10.57	both
NAD	15.01	both
NADH	14.45	both
NADP	17.62	both
NADPH	18.11	both
nicotinamide (niacinamide)	5.89	+
nicotinic acid	6.60	+
normetanephrine	11.60	+
Nα-acetylornithine	15.01	+
Nε-acetyllysine	12.32	+
ornithine	19.57	+
orotic acid	9.61	-
oxalic acid	18.46	-
oxaloacetate	17.88	-
pantothenate	7.67	both
para -coumaric acid	7.11	-
phenylalanine	8.93	+
phosphocreatine	16.07	both
phosphoenolpyruvate	18.43	both
proline	12.54	+
pyruvate	7.70	-
riboflavin	6.93	+
riboflavin 5'-phosphate	12.17	both
ribose 5-phosphate	16.59	-
ribulose 5-phosphate	16.37	-
S-adenosylhomocysteine	13.37	+
S-adenosylmethionine	15.87	+
sarcosine	13.50	+
sedoheptulose 7-phosphate	16.94	-
serine	15.56	both
serotonin	20.03	+
shikimate	14.42	-
sorbitol 6-phosphate	16.90	both
succinate	16.20	-
taurine	14.85	-
threonine	14.26	both
glucosamine 6-phosphate	11.48	both
glucose	7.48	-
glucose 1-phosphate	11.36	-
glucose 6-phosphate	11.51	-
glucuronic acid	10.42	-
glucuronic acid γ-lactone	3.57	both
glutathione, oxidized (GSSG)	11.73	both
glutathione, reduced (GSH)	10.73	both
GMP	12.58	both
GTP	15.22	both
guanine	8.08	both
guanosine	8.44	both
isocitrate	11.84	-
isomaltose	8.71	-
malate	10.75	-
maltose	8.61	-
mannitol 1-phosphate	10.87	both
mannose	7.26	-
mannose 6-phosphate	11.09	both
methylglyoxal	14.18	-
N6-methyladenosine	4.39	+
N6-methyl-AMP	11.32	both
N-acetylglucosamine	6.53	both
palmitoyl-CoA	11.32	+
pantothenate	9.03	-
phosphoenolpyruvate	13.02	-
Phosphoserine	11.73	both
Phosphothreonine	11.38	both
Phosphotyrosine	12.22	both
pseudouridine	7.43	-
riboflavin	5.20	+
riboflavin 5'-phosphate	11.50	both
ribulose 5-phosphate	10.83	-
sedoheptulose 7-phosphate	10.94	-
shikimate	10.33	-
succinate	10.62	-
sucrose	8.19	-
thymidine	3.33	-
thymine	3.08	-
TMP	11.38	both
UDP-glucose	11.23	-
UMP	11.20	both
uracil	4.08	-
xanthine	10.61	-
xanthosine	10.65	both
XMP	12.38	both
α-lactose	8.63	both
α-mannose 1-phosphate	10.98	-

thymine	6.87	-
tropic acid	6.18	-
tryptophan	10.95	+
tyramine	15.29	+
tyrosine	12.63	+
UDP-glucose	17.28	-
UMP	15.86	both
uracil	7.80	-
uric acid	12.70	-
valine	12.05	+
α-ketoglutarate	16.50	-
α-lipoic acid	4.67	-
α-tocopherol	3.70	-
β-alanine	14.99	+
γ-aminobutyric acid	15.20	+

DPYD is a key component of a metabolic program required for the epithelial-mesenchymal transition

Yoav D Shaul [1,2], Elizaveta Freinkman[1], William C Comb[1], Jason R Cantor[1], Wai Leong Tam[1], Prathapan Thiru[1], Dohoon Kim[1], Michael E Pacold[1,7], Walter W Chen[1], Brian Bierie[1], Richard Possemato[1], Robert A. Weinberg [1,4,6], Michael B Yaffe[2,4,5], David M Sabatini[1,2,3]

[1] Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology, Department of Biology, Nine Cambridge Center, Cambridge, MA 02142, USA

[2]Koch Institute for Integrative Cancer Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[3]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4]Departments of Biology and [5]Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[6]MIT Ludwig Center for Molecular Oncology, Cambridge, MA, USA

[7]Department of Radiation Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA, USA

*Correspondence: sabatini@wi.mit.edu

It is increasingly appreciated that oncogenic transformation alters cellular metabolism to facilitate cancer cell proliferation but less is known about the metabolic changes that promote cancer cell aggressiveness. Here, we analyzed the expression patterns of 1,704 metabolic genes in a large collection of cancer cell lines and found that the majority retained tissue-of-origin signatures. However, a set of high-grade carcinoma lines derived from diverse tissues shared a unique 44-gene signature, which we designate the "mesenchymal metabolic signature" (MMS) because these cells co-expressed a set of mesenchymal markers. In immortalized human mammary epithelial cells, a FACS-based shRNA screen identified several MMS genes as essential for these cells to undergo the epithelial-mesenchymal transition (EMT) but not for cell proliferation. Dihydropyrimidine dehydrogenase (DPYD), the rate-limiting enzyme for pyrimidine degradation, was highly expressed upon induction of the EMT program in mammary epithelial cells and its catalytic activity was found to be necessary for cells to acquire mesenchymal markers and to grow as mammospheres. Dihydrouracil, the immediate product of DPYD, also increased greatly upon EMT induction, and could substitute for DPYD in mammosphere formation. Thus, we identify metabolic processes, in particular pyrimidine degradation, as essential for the expression of the EMT program, a process associated with the acquisition of metastatic and aggressive cancer cell traits.

Introduction

Alterations in cellular metabolism are now recognized as an emerging hallmark of cancer (reviewed in (Hanahan and Weinberg, 2011; Hensley et al., 2013; Schulze and Harris, 2012)). Almost a century ago, Otto Warburg observed that, under aerobic conditions, tumor cells display increased glucose uptake and glycolytic rates compared to resting cells (reviewed in (Hsu and Sabatini, 2008; Vander Heiden et al., 2009; Ward and Thompson, 2012). Subsequently, many studies have revealed how this and other metabolic changes allow cancer cells to accumulate building blocks for the biosynthesis of macromolecules, while simultaneously maintaining energetic and redox balance (reviewed in (Cantor and Sabatini, 2012)). Whereas many of these mechanisms are shared with normal rapidly proliferating cells, in recent years cancer genomic data have revealed metabolic alterations that appear to occur only in specific tumor types. These changes include the loss of succinate dehydrogenase (SDH) or fumarate dehydrogenase (FH) in certain renal cell carcinomas and other familial cancer syndromes (reviewed in (Gottlieb and Tomlinson, 2005)), mutation of isocitrate dehydrogenase (IDH) 1 or 2 in glioma, acute myeloid leukemias, and chondrosarcomas (Amary et al., 2011; Mardis et al., 2009; Parsons et al., 2008), and amplification of phosphoglycerate dehydrogenase (PHGDH) in estrogen receptor (ER)-negative breast cancer and melanoma (Locasale et al., 2011; Possemato et al., 2011). These examples suggest that, in addition to fueling increased proliferation, cancer-associated alterations in metabolism can also satisfy tumor-specific demands.

Relatively few studies have examined the metabolic underpinnings of the cellular programs that increase cancer cell aggressiveness (Nomura et al., 2010; Ulanovskaya et al., 2013; Zhang et al., 2012). One such program is the epithelial-mesenchymal transition (EMT) (reviewed in (Nieto and Cano, 2012)) that operates in carcinoma cells and is thought to confer stem-like properties, such as enhanced survival, self-renewal, and anchorage-independent growth, all of which contribute to increased aggressiveness in vivo (Mani et al., 2008; Morel et al., 2008; Scheel and Weinberg, 2011). Indeed, EMT markers are predictive for

increased invasion, loss of differentiated characteristics, metastasis, and poor prognosis in a number of human tumor types (Nieto, 2011; Peinado et al., 2007; Singh and Settleman, 2010).

To understand how cellular metabolism contributes to these and other proliferation-independent features of cancer, we created a framework for the systematic identification of metabolic alterations specific to particular tumor types, as well as those that may characterize high-grade malignancies. By analyzing metabolic gene expression patterns in a large number of cancer cell lines, we identified a metabolic gene signature that is present in high-grade tumors bearing mesenchymal markers. Among the enzymes encoded by these genes is dihydropyrimidine dehydrogenase (DPYD), which catalyzes the rate-limiting step in pyrimidine degradation and whose physiological role in cancer was previously unknown. We find that EMT-promoting transcription factors induce the expression of DPYD and that its enzymatic activity is necessary for cancer cells to undergo an EMT. These findings reveal that the EMT induces a particular metabolic state and suggest that DPYD may have value as a diagnostic marker or therapeutic target in high-grade carcinomas.

Results

A mesenchymal-like metabolic gene expression signature in high-grade carcinoma cells

In order to study metabolic gene expression patterns in cancer, we used publicly available data to generate a database of mRNA expression profiles for 1,704 metabolic genes in 978 human cancer cell lines (see Methods) (Possemato et al., 2011). Aided by unsupervised hierarchical clustering, we organized the profiles into five distinct groups (Figure 1A); for four of these groups, the basis for clustering was readily apparent. One group consisted of melanoma cell lines, which uniquely express skin pigment biosynthesis genes. The cell lines in a second group were derived from hematopoietic system cancers (e.g., leukemia, lymphoma, and multiple myeloma) and, in a third, from neuroendocrine or neuroectodermal cancers (e.g., small cell lung cancer,

medulloblastoma, neuroblastoma (Onganer et al., 2005; Parham, 2001)). A fourth group consisted mostly of epithelial cancer cell lines, in which cell lines originating from breast, liver, colon, kidney, etc. clustered together. These results indicate that patterns of metabolic gene expression are sufficient to organize most cancer cell lines by tissue-of-origin, suggesting that many cancers retain significant portions of the metabolic programs of their normal tissue counterparts.

The cell lines in the fifth group proved more difficult to classify, and thus was initially named the “mixed-lineage group” (Figure S1A). While this group contained almost all the cell lines derived from mesenchymal tumors (soft tissue sarcoma, osteosarcoma; 20% of the cell lines in this group) and glioblastomas, it also included a large number of carcinoma lines (e.g. non-small-cell lung, hepatocellular, and breast; 43% of the cell lines in this group). Notably, all the breast cancer lines in the mixed-lineage group were of the Basal B subtype, which are derived from high-grade carcinomas (Carey et al., 2010) (Figure 1C). Likewise, all the hepatocellular carcinoma (HCC) cell lines in this group were also derived from high-grade tumors (Park et al., 1995), and retained fewer of the metabolic gene expression features of normal liver than did the HCC lines that were in the epithelial group (Figure S1B). Such loss of epithelial and gain of mesenchymal characteristics has been associated with high-grade malignancy in a variety of carcinoma types (Brabletz, 2012; Nieto, 2011). Moreover, several of the glioblastoma and the majority of Basal B breast cancer cell lines are known to bear mesenchymal characteristics (Kao et al., 2009; Neve et al., 2006; Verhaak et al., 2010). Thus, we thought it is likely that the cell lines in the mixed-lineage group shared a common mesenchymal-like phenotype.

Indeed, Gene Set Enrichment Analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) of ~17,000 genes showed that expression of the mesenchymal gene-set (EMT_UP) was significantly elevated in the mixed-lineage group relative to the other groups (FDR q-value<0.0001; Figure S1D). Furthermore, the mixed-lineage group had elevated expression of key mesenchymal markers (Mani et al., 2008; Nieto and Cano, 2012; Peinado et al., 2007), including vimentin (VIM), Snail family zinc finger 1 and 2 (SNAI1/2), N-

cadherin (CDH2), Twist basic helix-loop-helix transcription factor 1 (TWIST1), and the zinc finger E-box binding homeobox 1 (ZEB1) transcription factor (Figure 1D). Lastly, the epithelial markers claudin 1 (CLDN1) and E-cadherin (CDH1) were expressed at low levels in this group (Figure 1D). Collectively, these data suggested that the cell lines in the mixed-lineage group, regardless of tissue of origin, displayed a mesenchymal-like gene expression profile. Accordingly, we refer hereafter to the mixed-lineage group as the mesenchymal group of cell lines.

Identification of a mesenchymal metabolic gene expression signature

We identified a mesenchymal metabolic signature (MMS), composed of 44 metabolic genes associated with diverse metabolic pathways, as highly and differentially expressed in the mesenchymal group of cell lines relative to the other groups (see Methods) (Table 1 and Figure 2A). The MMS is particularly enriched for glycan biosynthesis genes (36% of the genes in the set), including glutamine-fructose-6-phosphate aminotransferase (GFPT2) and acetylhexosamine pyrophosphorylase (UAP1), which encode the rate-limiting and endpoint enzymes of the hexosamine biosynthetic pathway (HBP), respectively (Elbein et al., 2004; Zhang, 2004). The HBP end product, UDP-GlcNAc, is used by the enzyme O-GlcNac transferase (OGT) as a donor substrate to modify proteins via covalent attachment of GlcNAc to serine and/or threonine residues (Ma and Vosseller, 2013). Of special interest, this modification plays an important role in mesenchymal cells by stabilizing the EMT-inducing transcription factor SNAI1, which in turn down-regulates the key epithelial marker CDH1 (E-cadherin) (Park et al., 2010). The MMS list includes other genes with known connections to cancer aggressiveness, such as ecto-5'-nucleotidase (NT5E, also known as CD73), a mesenchymal stem cell marker (Lehmann et al., 2011; Zhi et al., 2012), and ectonucleotide pyrophosphatase/phosphodiesterase 2 (ENPP2), which promotes cell migration and metastasis (Ferry et al., 2008; Samadi et al., 2011). These examples suggest that the remaining MMS genes may also play an important role in the mesenchymal phenotype and/or aggressiveness of certain cancer cells.

We found that the MMS genes were significantly upregulated in cell lines that express known mesenchymal markers (Figure 2B, left). For example, this gene set is upregulated in cell lines derived from Basal B breast cancer and high-grade HCC relative to their luminal and low-grade counterparts, respectively (Figure 2C). Indeed, quantitative PCR and immunoblotting confirmed the overexpression of several individual MMS genes, including nicotinamide N-methyltransferase (NNMT) and DPYD, in high-grade breast cancer and HCC cell lines (Figure 2D and 2E), which also expressed mesenchymal markers, such as ZEB1 and TWIST1, and low levels of E-cadherin (CDH1) (Figure S2B).

We next asked if MMS gene expression correlates with that of known mesenchymal markers in primary human tumors as well as in cancer cell lines. From a database of expression profiles for 1,460 human primary tumors, including many of mesenchymal origin, we identified tumors with high expression of known mesenchymal markers (see methods) (Figure S2A). In such tumors, the MMS genes were significantly more highly expressed than in tumors not expressing these markers (Figure 2B and S2A). Thus, MMS gene expression correlates with that of known mesenchymal markers in both cancer cell lines and tumors, suggesting that a particular metabolic program characterizes the mesenchymal cell state.

EMT-dependent induction of mesenchymal metabolic signature genes

Given the high expression of MMS genes in mesenchymal-like relative to epithelial cancer cell lines, we hypothesized that the EMT program may directly affect the expression of these genes. To investigate this possibility, we examined engineered human mammary epithelial (HMLE) cells that undergo an EMT upon the activation of Twist (HMLE-Twist-ER) following treatment with hydroxytamoxifen (OHT) (Mani et al., 2008; Taube et al., 2010). Over a 15-day treatment with OHT, the HMLE-Twist-ER cells shifted their cell-surface markers from an epithelial ($CD24^{high}$, $CD44^{low}$) to a mesenchymal ($CD24^{low}$, $CD44^{high}$) profile (Al-Hajj et al., 2003), induced ZEB1 and TWIST1 expression, and suppressed CDH1 (E-cadherin) (Figure S2C) (Figure 2G and 2H). Like the mesenchymal markers, MMS genes such as DPYD and NNMT also displayed a

progressive increase in mRNA and protein levels over the course of OHT treatment (Figure 2G and 2H). Moreover, NAMEC cells, an HMLE-derived cell line that spontaneously acquired the mesenchymal state (Tam et al., 2013) (Figure S2D), also expressed high levels of several MMS genes (Figure 2H). Lastly, re-analysis of a previous expression profiling study comparing HMLE cells expressing an empty vector or Twist (Taube et al., 2010) showed that, unlike the majority of metabolic genes, MMS genes were upregulated upon EMT induction in culture (Figure 2F). Collectively, these results suggest that the EMT program and MMS gene induction are coupled processes.

A FACS-based pooled shRNA screen for MMS genes required for the EMT

To identify which, if any, of the MMS genes play a critical role in the EMT, we developed a FACS-based RNAi screen using a pool of 514 lentivirus vector-expressed shRNAs targeting 42 of the MMS genes, as well as known control genes (GFP, RFP, Luciferase, and LacZ), and non-MMS metabolic genes (Figure S3A). We then induced the EMT in HMLE-Twist-ER cells expressing the shRNA hairpin library, and after 15 days compared the abundance of each hairpin in FACS-sorted epithelial and mesenchymal cell populations isolated using the CD44 and CD24 surface antigens (Figure 3A). We reasoned that knockdown of an EMT-essential gene would cause cells to remain in the epithelial state ($CD24^{high}CD44^{low}$) even upon OHT treatment. Indeed, hairpins targeting the EMT-promoting transcription factors ZEB1 and SNAI1 were enriched in the epithelial population (Figure 3B).

We also found that hairpins against 16 MMS genes were similarly enriched, suggesting that knockdown of these genes blocks activation of the EMT program (Figure 3C). Among the MMS genes, *DPYD* was a top hit, with 5 out of 12 hairpins scoring in the screen (Figure 3B). DPYD is the rate-limiting enzyme of the pyrimidine degradation pathway (Amstutz et al., 2011) and is also capable of degrading the chemotherapeutic agent 5-fluorouracil (5-FU), but the physiological role of this enzyme in cancer cells is unclear (Amstutz et al., 2011; Mizutani et al., 2003; Offer et al., 2013; Yoo et al., 2009).

We wished to rule out the possibility that knockdown of DPYD and the other MMS hit genes may block the EMT by affecting the proliferation or viability of epithelial cells. Thus, in a parallel experiment, we determined the abundance of each hairpin in HMLE-Twist-ER cells before and after a 15-day period of proliferation in the absence of EMT induction (Figure 3A, uninduced day 0 and day 15). As expected, the control hairpins as a group had a neutral effect on proliferation (median log2 hairpin abundance ratio = -0.28). Importantly, the abundance distributions of the ZEB1, SNAI1, and DPYD hairpins did not differ significantly from the control group (Figure 3D), indicating that these hairpins did not affect cellular viability or proliferation; by contrast, hairpins targeting ribonucleotide reductase M1 (RRM1) and thymidylate synthetase (TYMS), which are critical for cell division (Tennant et al., 2010), caused a significant anti-proliferative effect (median log2 hairpin abundance ratio = -3.23, and -2.4, respectively) (Figure 3D). Therefore, knockdown of DPYD suppressed the EMT program without inhibiting the viability or proliferation of epithelial cells, suggesting that this enzyme plays a specific role in inducing the mesenchymal cell state.

DPYD expression promotes the EMT

To further establish the role of DPYD in the EMT, we individually infected HMLE-Twist-ER cultures with eight distinct shRNAs targeting DPYD, and found that DPYD knockdown, in a dose-dependent manner, decreased the percentage of cells with a mesenchymal profile ($CD24^{low}/CD44^{high}$) after 15 days of Twist induction by OHT treatment (Figure S4A). In addition, DPYD knockdown with the hairpins (shDPYD_1 and shDPYD_4) that most strongly reduced DPYD expression (Figure S4B), did not affect the viability of untreated HMLE-Twist-ER cells (Figure S4C), but decreased the percentage of OHT-treated cells with a mesenchymal profile (Figure 4A) and suppressed the expression of ZEB1 and VIM (Figure 4B and S4D). Moreover, DPYD knockdown also decreased the capacity of the cells to form mammospheres, a unique property of the mesenchymal-like ($CD24^{low}/CD44^{high}$) but not epithelial ($CD24^{high}/CD44^{low}$) HMLE cells (Dontu, 2003; Mani et al., 2008) (Figure 4C). Thus, this functional assay

confirmed that a reduction in DPYD expression inhibited expression of the EMT program.

To rule out the possibility that the effects of the DPYD shRNAs are due to off-target effects, we restored DPYD levels in shDPYD-expressing HMLE-Twist-ER cells by ectopically expressing the mouse isoform of DPYD (mDPYD), which is 86% identical at the amino acid level to the human isoform but unaffected by the shRNAs targeting human DPYD (Figure S4E). Expression of mDPYD in the presence of shDPYD_1 fully restored EMT induction to the level observed in the shGFP control (Figure 4D). Additionally, we observed that in the mDPYD-rescued cells, the expression of the mesenchymal markers ZEB1 and VIM (Figure 4E) and the capacity for mammosphere formation (Figure 4F) were also restored. Interestingly, we noted that ectopic expression of mDPYD could further increase the percentage of mesenchymal-like cells over that of the empty-vector control (Figure 4D, compare the top left and bottom left panels), suggesting that the expression level of DPYD is a limiting factor in activating the EMT program. Thus, we conclude that DPYD expression is elevated during the EMT program and plays an essential role in this process.

Cellular dihydropyrimidine levels are elevated during the EMT

Having demonstrated that DPYD expression plays a critical role in the EMT program, we asked whether its metabolic products increase in abundance during this process. To do so, we used liquid chromatography and mass spectrometry (LC-MS) (Büchel et al., 2012) to determine the cellular concentration of DPYD substrates (uracil and thymine) and immediate products (dihydrouracil and dihydrothymine) (Figure 5A) (Lohkamp et al., 2010). In HMLE-Twist-ER cells, overexpression or knockdown of DPYD resulted in a corresponding ~10-fold increase or decrease, respectively, in the intracellular DHU/uracil molar ratio (Figure 5B). Moreover, NAMEC cells exhibited higher DHU/uracil and DHT/thymine ratios than HMLE-Twist-ER cells (by 10- and 6-fold, respectively; Figure 5B and S5A), consistent with the higher endogenous DPYD expression level in the former cells (Figure 2H). In addition, OHT treatment of HMLE-Twist-ER cells, which progressively upregulates DPYD expression (Figure

2G and 2H) gradually increased the cellular DHU/uracil molar ratio by 5-fold after 15 days of Twist induction (Figure 5C). DPYD expression and DHU/uracil ratios were also correlated in breast cancer and HCC cell lines (Figure 5D and 5E). Notably, the higher DHU/uracil molar ratio in MCF7 breast cancer cells compared to the other luminal breast cancer cell lines (Figure 5D) correlated with the relatively high expression of DPYD in this particular cell line (Figure 2D). Hence, DHU/uracil ratios correlate closely with DPYD expression levels and mesenchymal character in a number of cellular settings, suggesting that DPYD is enzymatically active in the cancer cell lines that we examined.

DPYD is normally expressed in the liver, where it is the rate-limiting enzyme of a three-step pyrimidine degradation pathway that converts uracil and thymine to β-alanine and 2-methyl-β-alanine, respectively (Figure 5A) (Lohkamp et al., 2010). In the liver, the immediate products of DPYD are further catabolized by dihydropyrimidinase (DPYS) and beta-ureidopropionase (UPB1) (Amstutz et al., 2011; Van Gennip and Van Kuilenburg, 2000) (Figure 5A). By contrast, we found that HMLE-Twist-ER and NAMEC cells express only DPYD, but not the other components of this catabolic pathway (Figure S5B). In addition, unlike DPYD expression, DPYS and UPB1 expression was not elevated in breast Basal B and high-grade HCC cell lines (Figure S5C). These observations may explain why the products of DPYD activity accumulate in mesenchymal-like cancer cells, but not in normal liver (Lohkamp et al., 2010).

DPYD activity is essential for its function in the EMT

The accumulation of DPYD products in mesenchymal-like cells suggests that its function in the EMT is mediated through its enzymatic activity. Accordingly, we tested the ability of a catalytically attenuated mouse DPYD mutant (mDPYD-I560S, also known as DPYD*13, which has a 75% reduction in enzymatic activity relative to WT (Ezzeldin and Diasio, 2004; Offer et al., 2013)), to rescue the inhibitory effect of shDPYD_1 on EMT induction. Whereas expression of wild-type mDPYD in the presence of shDPYD_1 restored the EMT induction following 15 days of OHT treatment, mDPYD-I560S had a greatly reduced capacity to rescue CD44/CD24 expression and mammosphere

formation, and completely failed to restore expression of the EMT-inducing transcription factor ZEB1 (Figure 6A-C). In addition, we found that, while control cells (expressing empty vector) treated with OHT for only 10 days displayed an intermediate CD44/CD24 marker expression profile, cell lines ectopically expressing either mouse or human DPYD (DPYD-FLAG) displayed higher mesenchymal marker expression at this earlier time point (Figure 6D), resembling the profile of control cells after a full 15 days of OHT treatment (Figure 6A). In contrast to wild-type DPYD, overexpression of the mutant DPYD-I560S (human DPYD-I560S-FLAG) had a greatly attenuated effect on cell-surface marker expression and mammosphere formation, while preventing ZEB1 expression (Figure 6D-F). Thus, the physiological role of DPYD in the EMT program requires its enzymatic activity. Moreover, the accelerated progression of the EMT in DPYD-overexpressing cells suggests that DPYD products may be rate-limiting in this process.

To further establish the role of DPYD products in the EMT, we asked whether addition of DHU or DHT to culture media could substitute for DPYD loss. Indeed, treatment of shDPYD_1 cells with these metabolites at 10 or 100μM resulted in a dose-dependent rescue of mammosphere formation (Figure 6G), whereas the DPYD substrate uracil had a significantly smaller effect (Figure S6A), despite the fact that uracil and DHU accumulated to comparable intracellular concentrations (data not shown). Therefore, the effect of DPYD knockdown on mammosphere formation can be reversed either by ectopic expression of active DPYD (Figure 4F and 6C) or by supplementation of its products to the cell culture media. Together, these results confirm that the MMS gene product DPYD plays a critical role in the EMT via its enzymatic activity and dihydropyrimidine production.

Discussion

We identified a mesenchymal metabolic signature (MMS) consisting of 44 metabolic genes upregulated in cancers bearing mesenchymal markers. Several of these metabolic genes are essential for the EMT, including DPYD, the rate-

limiting enzyme of the pyrimidine degradation pathway. Remarkably, the expression of DPYD is not essential for cell viability or proliferation, demonstrating the existence of metabolic processes that specifically enable carcinoma cells to acquire mesenchymal-like characteristics. Because these characteristics are associated with increased cancer aggressiveness, these findings suggest that DPYD activity may play a role in carcinoma progression.

There is a clear difference between the metabolic pathways that are associated with proliferation and those upregulated during the EMT. Compared to resting cells, proliferating cells upregulate glycolysis and nucleotide biosynthesis pathways (Hu et al., 2013), whereas the mesenchymal metabolic signature (MMS) is enriched for glycan biosynthesis genes. Glycosylation is thought to be one of the most common covalent protein modifications in eukaryotic cells, with a major role in differentiation and mediating cell-cell interactions (Li et al., 2013). Because the EMT is accompanied by major changes in cell morphology and detachment from the surrounding cells, it is reasonable to assume that a major glycan remodeling may occur during the EMT program. Furthermore, glycosylation regulates the function of several key players in the EMT, including the products of the SNAI1 and CD44 genes (Jaggupilli and Elkord, 2012; Park et al., 2010). Thus, we anticipate that future studies will further demonstrate an important role for specific glycan remodeling events in both the mesenchymal phenotype and in the EMT program.

After executing the EMT program, epithelial-derived cancer cells acquire traits associated with high-grade malignancy, including resistance to apoptosis and chemotherapy, dedifferentiation, and invasiveness, which can lead to metastatic dissemination from primary tumors (Brabletz, 2012; Nieto and Cano, 2012; Scheel and Weinberg, 2011). Thus, inhibiting the EMT may maintain a tumor in a lower-grade state, potentially increasing therapeutic efficacy and slowing metastasis. The feasibility of manipulating epithelial plasticity is reinforced by studies showing that depletion of ZEB1 by RNA interference in mesenchymal-like cells results in a partial mesenchymal-epithelial transition (MET), presumably through the induction of CDH1 (E-cadherin) expression

(Aigner et al., 2007; Chaffer et al., 2013; Li et al., 2009). However, the development of inhibitors targeting transcription factors such as ZEB1 remains a challenge (Singh and Settleman, 2010). By contrast, many of the enzymes encoded by the MMS have well-defined active sites that can potentially be targeted by small molecules. Here, we demonstrate that DPYD expression and activity are essential for the induction of ZEB1 expression, suggesting that the expression of transcriptional drivers of the EMT program can be modulated through inhibition of metabolic enzymes such as DPYD.

Many studies have linked DPYD function with acquired tumor resistance to the chemotherapeutic agent 5-FU, but the physiologic role of this enzyme in cancer cells is unknown (Amstutz et al., 2011; Mizutani et al., 2003; Offer et al., 2013; Yoo et al., 2009). By demonstrating that DPYD plays an essential role in the EMT, we provide one of the first indications for its function in cancer. However, there is a clear distinction between this function and the normal role of DPYD in the liver. In the latter, DPYD functions as the first enzyme in a three-step pathway of pyrimidine degradation, whereas we show that in mesenchymal-like cells, the expression level of the two downstream enzymes (DPYS and UPB1) is not detectable at the mRNA level. Therefore, the EMT program reconfigures the pyrimidine degradation pathway in order to use only DPYD, presumably because its enzymatic activity fulfills a specific metabolic demand. We suggest that this EMT-dependent metabolic rewiring, which activates only selected components of a given metabolic pathway, is not exclusive to DPYD, but can potentially occur in other MMS-related metabolic processes. Thus, through such rewiring, the EMT may confer novel cellular functions to other pathways represented in the MMS as well. Further studies aimed at understanding the role of the MMS genes in cancer may reveal novel metabolic processes that promote cancer aggressiveness.

The function of DPYD in the EMT is dependent upon its products, the dihydropyrimidines (DHPs), DHU and DHT. However, understanding the role of these metabolites in the EMT program is challenging, because no biological function has been reported for the DHPs other than as substrates for the enzyme

DPYS. One possibility is that DHPs may act as allosteric regulators of other enzymes, similar to serine in the case of the glycolytic enzyme pyruvate kinase M2 (PKM) (Chaneton et al., 2012), or as receptor ligands, like the citric acid cycle intermediates succinate and α-ketoglutarate in the case of the G-protein coupled receptors GPR99 and GPR91, respectively (He et al., 2004). In this scenario, the DHPs themselves could act as key signaling molecules without further enzymatic processing.

Another potential function for the DHPs is that these pyrimidine bases could be converted to pyrimidine deoxynucleosides or nucleosides and thus possibly incorporated into DNA or RNA, respectively. Support for this latter possibility comes from previous studies showing that genotoxic agents can damage DNA precursors (dNTPs) in the nucleotide pools of bacterial cells (Dolinnaya et al., 2013). These chemically altered dNTPs, including the deoxynucleotide triphosphate form of DHT (DHdTTP), have been found to be incorporated into bacterial genomes (Dolinnaya et al., 2013) and are able to substitute for deoxythymidine triphosphate (dTTP) as substrates for the *E. coli* DNA polymerase I and Klenow fragments (H Ide, 1988; Ide et al., 1987). It remains to be determined whether such modified nucleotides can be produced in human cells and, if so, how they affect cellular phenotypes.

Materials and methods

Antibodies

Antibodies were obtained from the following sources: Epithelial-Mesenchymal Transition (EMT) Antibody Sampler Kit (89782) (includes antibodies for ZEB1, VIM, CDH1, and SLUG), DPYD (4654), and Actin (3700) from Cell Signaling Technology; FITC-labeled anti-CD24 (555427), and APC-labeled anti-CD44 (559942) from BD Bioscience; HRP-labeled anti-mouse and anti-rabbit secondary antibodies from Santa Cruz Biotechnology.

Cell Lines and Cell Culture

The immortalized human mammary epithelial cells expressing ectopic Twist-ER (HMLE-Twist-ER) and Naturally Arising MEsenchymal Cells (NAMECs) have been described ((Elenbaas, 2001; Mani et al., 2008) and (Tam et al., 2013), respectively. HMLE-Twist-ER and NAMEC cells were maintained in MEGM (Lonza) growth media. The cell lines ZR-75-1, EVSA-T, MCF7, MDA-MB-231, MDA-MB-157, Hs-578-T, HEPG2, SNU-387, and SNU-432 were obtained from ATCC and were maintained in DMEM supplemented with 10% IFS. All cells were cultured at 37°C with 5% CO_2. For EMT induction, HMLE-Twist-ER cells were treated with 4-hydroxytamoxifen (OHT) (Sigma, H7904) at a final concentration of 10nM for the indicated number of days.

Cancer cell line gene expression matrix and median of median determination

Cancer cell line gene expression data were collected from (1) the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), (2) GlaxoSmithKline (GSK) cell line data (https://cabig.nci.nih.gov/caArray_GSKdata/), (3) and Gene Expression Omnibus database (GEO) (Barrett et al., 2007). Data were normalized by RMA using the Affymetrix package from Bioconductor. A custom probeset definition was used for processing the arrays as defined by Dai M et al (Dai et al., 2005) such that there was one probeset per Entrez Gene ID. The cell lines were classified based on their tissue of origin (with the exception of breast and lung cell lines, which were further divided based on Estrogen Receptor status (for breast) or SCLC and NSCLC (lung)), resulting in 22 different groups. In order to avoid bias toward tissues that are represented by a large number of cell lines, we calculated the cancer cell lines median in two steps. First, the median expression value for each gene among the cancer cell lines from a single tissue of origin was calculated, resulting in one value for each gene in each tissue of origin. Second, these tissue-of-origin median values were combined, and their median was determined to obtain the “cancer cell line median of medians” value for each gene. The relative gene expression level for each

metabolic gene in each cell line was calculated as the ratio of its expression level to the corresponding median of median value (Table S1).

Primary tumor gene expression matrix and median of median determination

Primary tumor gene expression data were collected from (1) “Expression Project for Oncology” (http://www.intgen.org/expo/) and (2) Gene Expression Omnibus database (GEO) (Barrett et al., 2007). Data were normalized by RMA using the Affymetrix package from Bioconductor. A custom probeset definition was used for processing the arrays as defined by Dai M et al (Dai et al., 2005) such that there was one probeset per Entrez Gene ID. The calculation for the primary tumor median of medians was conducted similarly to that of cancer cell lines median of medians.

Identification of the Metabolic Mesenchymal Signature (MMS) genes

For each metabolic gene, the ratio between the mean expression level in mesenchymal (mesenchymal group, Figure 1) and non-mesenchymal cell lines (all other groups) was determined. The mean and standard deviation of all the metabolic gene expression ratios was calculated, and all genes upregulated above a Z-score of 2.5 or below a Z-score of -2 were classified as MMS (Table S2).

Fluorescence-Activated Cell Sorting (FACS) Analysis

Cells were prepared according to standard protocols and suspended in 1% Serum/PBS on ice prior to FACS. 7-AAD (Life Technologies) was used to exclude dead cells. Cells were sorted on a BD FACSAria or analyzed using the FACSCalibur HTS (BD Biosciences) with FlowJo software (Tree Star, Ashland, OR).

RNA Preparation and RT-PCR Analysis

Total RNA was isolated from cells or tissues using the RNeasy Kit (Qiagen, 74106) and reverse-transcription was performed using Superscript III reverse transcriptase (Invitrogen, 18080-044). The resulting cDNA was diluted in DNase-

free water (1:10) before quantification by real-time quantitative PCR. mRNA transcription levels were measured using SYBR Green PCR master mix (Applied Biosystems, 430955) and Biosystems 7900HT sequence Detection System v2.3 software. All data are expressed as the ratio between the expression level of the target gene mRNA and that for Actin. Primers used for qRT-PCR were obtained from Integrated DNA Technology and are listed in the table below. Human adult liver total RNA was from Cell Application (1H21-50).

Primers used for qRT-PCR

Genes	**Forward**	**Reverse**
CYBRD1	TCGTCTGGGTCCTCCACTAC	TGGCAGCAACTGCATTTAAC
DPYD	GTGTTCCACTTCGGCCAAGAA	GAGTCGTGTGCTTGATGTCAT
DSE	GGGCTCCAGTGTGTTTTTCA	GTCGGTGATGTAGGCTGACA
DSEL	GGCCTTGGTGACTGGAGTAG	GCTGGGCCAGAAAAACATAC
GPX8	ACTTCAGCGTGTTGGCTTTT	AGGCCTGATGACTTCAATGG
GXYLT2	GCTTGGGAGGACATGTTGTA	CAGTGATCGGGACGGTAGTT
HS3ST3A1	TGGAGAAGACGCCCAGTTAC	GACAGCGTCTGCGTGTAGTC
MME	AGAAGAAACAGCGATGGACTCC	CATAGAGTGCGATCATTGTCACA
NNMT	GACATCGGCTCTGGCCCCACT	GACATCGGCTCTGGCCCCACT
PPAP2B	TGAGAGCATCAAGTACCCACT	ACGTAGGGGTTCTGAATCGTC
HAS2	CTCTTTTGGACTGTATGGTGCC	AGGGTAGGTTAGCCTTTTCACA
ZEB1	TGCACTGAGTGTGGAAAAGC	TGGTGATGCTGAAAGAGACG
CDH1	TTGCACCGGTCGACAAAGGAC	TGGATTCCAGAAACGGAGGCC
VIM	ACCCGCACCAACGAGAAGGT	ATTCTGCTGCTCCAGGAAGCG
DPYS	ATTGATTTCGCCATTCCTCAGAA	GCTGTAGTCGCAGCAAACTTT
UPB1	GCGCGTTCTCTATGGCAAG	CCGCTGCTTCAAAGGCATATC
TWIST	TGCGGAAGATCATCCCCACG	GCTGCAGCTTGCCATCTTGGA

Pooled shRNA screen

pLKO.1 lentiviral plasmids encoding shRNAs targeting 74 genes (listed in Table S3)) were obtained and combined to generate a plasmid pool (Possemato et al., 2011). HMLE cells were infected with the pooled lentivirus at an MOI of 0.2-0.5 so as to ensure that most cells contained only one viral integrant. Cells were selected for 3 days with 0.5 mg/ml puromycin, after which 10^6 cells were removed, washed, and frozen at -80˚C (Figure 3A, day 0). The remaining cells were split into OHT-treated and untreated samples. After 15 days, the OHT-treated cells were trypsinized, washed with phosphate buffered saline (PBS)+1% inactivated fetal calf (IFC) serum, and FACS-sorted using CD44/CD24 antibodies in order to separate the mesenchymal and epithelial populations.

Genomic DNA was isolated from all the cells using the QIAampDNA mini kit (Qiagen). To amplify the shRNAs encoded in the genomic DNA, PCR was performed for 33 cycles at an annealing temperature of 66˚C using 3.5 µg of genomic DNA, the primer pair indicated below, and DNA polymerase (TAKARA Ex taq, Clontech lab). Forward primers containing unique 4-nucleotide barcodes were used (see below) so that PCR products obtained from many samples could be sequenced together. After purification, the PCR products from each cell sample were quantified by ethidium bromide staining (Sigma) after gel electrophoresis, pooled in equal proportions, and analyzed by high-throughput sequencing (Illumina). The shRNAs from all 4 DNA samples (day 0, day 15 untreated, day 15 OHT-treated mesenchymal, and day 15 OHT-treated epithelial) were sequenced together. Sequencing reads were de-convoluted using GNU Octave software by segregating the sequencing data by barcode and matching the shRNA stem sequences to those expected to be present in the shRNA pool, allowing for mismatches of up to 3 nucleotides. The $\log_2$ values reported are the average $\log_2$ of the fold change in the abundance of each shRNA in the mesenchymal-like samples compared to epithelial cells. The mean and standard deviation of the control hairpins (GFP, RFP, Luciferase, LacZ) were calculated and used to set a cutoff (one standard deviation below the control mean). Every

gene that had at least two hairpins with a $\log_2$ value below the cutoff was considered a hit (hairpin ratio list is in Table S4).

Primers for deep sequencing deconvolution

Forward PCR primers (Ns indicate location of 4nt barcode):	AATGATACGGCGACCACCGAGAAAGTATTTCGATT TCTTGGCTTTATATATCTTGTGGANNNNACGA
Reverse PCR primer:	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTT GTGGATGAATACTGCCATTTGTCTCGAGGTC
Sequencing Primer:	GAGAAAGTATTTCGATTTCTTGGCTTTATATATCTT GTGGA

Mammosphere Assay

500 cells/well were seeded in 96-well ultra-low adhesion plates (Corning, 3474) in MammoCult Basal Medium (Stem Cell Technology, 05621) containing 2.6% methylcellulose (Stem Cell Technology, H4100) and 10% MammoCult Proliferation Supplements (Stem Cell Technology, 05621), supplemented with 0.5µg/ml hydrocortisone, 4µg/ml Heparin and Pen/strep. Spheres were counted 12-14 days later.

Metabolite extraction

Solvents were obtained from Fisher Scientific and were Optima LC/MS grade except where otherwise specified. Cells grown in standard tissue culture plates (~500,000 cells per sample) were washed twice in an ice-cold solution of 0.9% NaCl in deionized water, followed by extraction on dry ice in 1 mL 80% methanol containing 10ng/mL phenylalanine-d_8 and valine-d_8 (Sigma-Aldrich) as internal standards. The cell mixtures were shaken vigorously on a Vortex mixer for 10 min. at 4°C, vacuum-dried, and resuspended in 100µL LC/MS grade water (Fisher). These extracts were then centrifuged at 15,000xg at 4°C for 10 min., and the supernatants were passed through a cellulose acetate particulate filter (National Scientific).

Liquid Chromatography (LC) analysis

An UltiMate 3000 UPLC system with autosampler (Dionex) was used for this study. Biological triplicate samples (typically 10µL) were injected onto an Atlantis dC18 2.1 x 150mm (3µm particle size) column (Waters) and eluted isocratically in a mobile phase consisting of 1mM ammonium acetate, 5mM formic acid, and 3.3% methanol (mobile phase A) at a flow rate of 0.2mL/min. The run time was 19 min.; the autosampler was held at 4°C and the column compartment was held at 12.5°C. To minimize carryover, blank injections were performed after every six analytical runs. In addition, after every 12 analytical runs, the column was cleaned with a gradient from 100% mobile phase A to 100% acetonitrile over 10 min., followed by 15 min. at 100% acetonitrile, and finally by 15 min. re-equilibration in 100% mobile phase A, all at 0.2mL/min.

Mass Spectrometry (MS) analysis

The UPLC system was coupled to a QExactive orbitrap mass spectrometer equipped with a HESI II probe (Thermo Fisher Scientific) operating in positive ion mode. The spray voltage was set to 3.9 kV, and the heated capillary and the HESI probe were both held at 270°C. The sheath gas flow was set to 28 units, the auxiliary gas flow was set to 13 units, and the sweep gas flow was set to 5 units. External mass calibration was performed every 7 days. The MS data acquisition was performed by targeted Selected Ion Monitoring (tSIM) of the metabolites of interest and the internal standards, with the resolution set at 35,000, the AGC target at 10^5, the maximum injection time at 250msec, and the isolation window at 1.0m/z. The full scan range was 70-1000 m/z. Quantitation of the data was performed with XCalibur QuanBrowser 2.2 (Thermo Fisher Scientific) using a 5 ppm mass tolerance, by a researcher blinded to the identity of the samples. Pure thymine (T0376) and uracil (U1128) were obtained from Sigma-Aldrich, and dihydrothymine (L01996) and dihydrouracil (L01918) were obtained from Alfa Aesar, and were run in half-log serial dilution (3nM -100µM) to confirm chromatographic retention times and generate standard curves for quantitation of each analytical batch.

Figure legends

Figure 1: Based on metabolic gene expression patterns, high-grade carcinoma cell lines co-cluster with mesenchymal cells

(A) Metabolic gene expression patterns are sufficient to cluster most, but not all, cancer cell lines based on their tissue of origin. Two-way hierarchical clustering of the expression levels of 1,704 metabolic genes in 978 different cell lines is presented as a heatmap. The clustering segregates cells into five groups that are named based on their common tissue of origin and are marked with a colored dendrogram. Values represent the $\log_2$ ratio of each expression level to the cancer cell line median of medians (see methods).

(B) Cell lines derived from related cancer types co-cluster based on metabolic gene expression patterns. Each row shows all the cell lines in the dataset derived from the indicated cancer type. Within each row, each black line represents the position of a cell line in the cluster.

(C) Many high-grade hepatocellular carcinoma (HCC) and basal B breast cancer cell lines mostly cluster within the mesenchymal group. The HCC and breast cancer cell line distributions are presented as in (B).

(D) Known mesenchymal markers are highly expressed in the mesenchymal group. Cancer cell lines were ordered identically as in (A). The heatmap represents the expression of known mesenchymal and epithelial markers in each cancer cell line. Values represent the $\log_2$ ratio of each expression level to the cancer cell line median of medians. Color bar shows Log_2 scale.

Figure 2: High expression of mesenchymal metabolic signature (MMS) genes in mesenchymal cell lines

(A) Identification of the MMS. For each metabolic gene, the ratio between the mean expression level in the mesenchymal group of cell lines and in all other groups (see Figure 1) was determined and used to rank the genes. The plot displays the distribution of the gene expression $\log_2$ ratio (y axis)

vs. gene rank (x axis). Genes that are upregulated (purple, 44 genes) or downregulated (blue, 16 genes) by at least 2-fold in mesenchymal relative to non-mesenchymal cells are highlighted.

(B) Elevated MMS gene expression in mesenchymal cancer cell lines and primary tumors. Cancer cell lines and primary tumors were divided into mesenchymal and non-mesenchymal groups based on the expression of known mesenchymal markers (Figure 1D and Figure S2A). For each metabolic gene, the ratio of the mean expression level between the groups was determined. The box plots represent the $\log_2$ ratio distribution of MMS genes (purple) and all other metabolic genes (gray). The P values for the indicated comparisons were determined using Student's T test.

(C) MMS gene expression is elevated in Basal B breast and high-grade HCC cancer cell lines. Box plots represent the expression levels of the MMS genes in the indicated breast cancer (green, luminal; red, Basal B) and HCC (blue, low-grade; brown, high-grade) subtypes. The P values for the indicated comparisons were determined using Student's T test.

(D) Individual validation of MMS mRNA levels in breast cancer (green, luminal; red, Basal B) and HCC (blue, low-grade; brown, high-grade) cell lines by quantitative real-time PCR (qPCR). Each value represents the mean ± SEM for n=3.

(E) Individual validation of MMS protein levels in the indicated breast cancer and HCC cell lines by immunoblotting.

(F) MMS genes are upregulated during the EMT. Microarray analysis for gene expression changes during EMT was described previously ((Taube et al., 2010), GSE24202). Here the same dataset was reanalyzed for the $\log_2$ expression ratio of MMS and all other metabolic genes between HMLE-Twist-ER cells forced to express Twist and Snai1 (mesenchymal) to HMLE-Twist-ER expressing empty vector (epithelial). The box plots represent the $\log_2$ ratio expression distributions of MMS genes (purple) and all other metabolic genes (gray). The P value for the comparison between the two groups was determined using Student's T test.

(G) MMS gene upregulation in an HMLE-Twist-ER inducible EMT system. HMLE-Twist-ER cells were treated with hydroxytamoxifen (OHT) to induce an EMT for 15 days. Every three days, cells were collected and mRNA isolated and subjected to qPCR using the indicated probes. Each value represents the mean ± SEM for n=3.

(H) MMS protein upregulation in the same cells as in (G). Every three days, cellular proteins were isolated and subjected to immunoblotting using the indicated antibodies. NAMEC cells are mesenchymal cells derived from HMLE cells (see methods).

Figure 3: A FACS-based pooled shRNA screen identifies DPYD as required for EMT

(A) Schematic presentation of the FACS-based pooled shRNA screen. OHT, hydroxytamoxifen; gDNA, genomic DNA.

(B) DPYD knockdown inhibits the EMT. All hairpins from the screen were ranked based on the $\log_2$ ratio of their abundance in the epithelial relative to the mesenchymal population of OHT-induced HMLE-Twist-ER cells after FACS sorting (see Figure 3A). Hairpin sub-pools pictured include those targeting control genes (39 hairpins targeting RFP, GFP, luciferase, and LacZ), ZEB1 (9 hairpins), SNAI1 (8 hairpins), and DPYD (12 hairpins). One standard deviation below the mean of the distribution of the control hairpins was set as a cutoff (red line). Every hairpin with a $\log_2$ ratio below the cutoff was considered a hit.

(C) Several of the MMS genes are critical for the EMT. Genes with at least two hairpins scoring below the cutoff (see panel B) were classified as hit genes. The numbers in the table represent the hit genes as a fraction of the total genes in a given sub-pool.

(D) DPYD knockdown does not affect cell viability. All hairpins were ranked based on the $\log_2$ ratio of their abundance in uninduced HMLE-Twist-ER cells on day 15 relative to day 0. The same hairpin sub-pools as in (B), with the addition of shRNAs targeting the essential genes RRM1 (4 hairpins) and TYMS (5 hairpins), are shown. The significance of the

differences in distribution between the selected genes and the control genes was quantified using Student's T test.

Figure 4: DPYD expression is essential for EMT induction

(A) DPYD knockdown (KD) inhibits the EMT. HMLE-Twist-ER cells were infected with hairpins against GFP (shGFP) and DPYD (shDPYD_1, shDPYD_4). The cells were either left untreated or treated with OHT for 15 days, as indicated, followed by FACS analysis of the cell-surface markers CD24 and CD44 to separate the epithelial and mesenchymal populations. The percentage of cells in each gate is presented.

(B) DPYD KD down-regulates ZEB1 expression. Cells infected with the indicated hairpins were treated with OHT for 15 days and subjected to immunoblotting with the corresponding antibodies.

(C) Quantification of in vitro mammosphere formation by cells treated as in (A). The data are reported as the number of mammospheres formed per 500 seeded cells; each value represents the mean ± SD for n=6. The P values for the indicated comparisons were determined using Student's T test.

(D) Mouse DPYD expression rescues the effects of DPYD KD on the EMT. HMLE-Twist-ER cells were infected with virus not expressing a cDNA (empty vector) or expressing mouse DPYD (mDPYD), together with either shGFP or shDPYD_1. The cells were either left untreated or treated with OHT for 15 days, as indicated, followed by FACS analysis of the cell-surface markers CD24 and CD44. The percentage of cells in each gate is presented.

(E) Mouse DPYD rescues the effects of DPYD KD on ZEB1 expression. HMLE-Twist-ER cells infected with the indicated hairpins and vectors were either left untreated or treated with OHT, followed by immunoblotting with the indicated antibodies.

(F) Mouse DPYD rescues the effects of DPYD KD on mammosphere formation. Quantification of in vitro mammosphere formation by cells treated as in (D). The data are reported as the number of mammospheres

formed per 500 seeded cells; each value represents the mean ± SD for n=6. The P value measured between the indicated samples was quantified using Student's T test.

Figure 5: The products of DPYD are elevated in mesenchymal cells

(A) Schematic presentation of the pyrimidine degradation pathway. Gene names are marked in red: DPYD, dihydropyrimidine dehydrogenase (rate-limiting step); DPYS, dihydropyrimidinase; UPB1, beta-ureidopropionase. Metabolites: DHU, dihydrouracil; DHT, dihydrothymine.

(B) Modulation of DPYD expression affects the cellular DHU/uracil molar ratio. DHU and uracil levels were measured by liquid chromatography and mass spectrometry (LC-MS) in NAMEC or HMLE-Twist-ER cell lines expressing empty vector, DPYD-FLAG or shDPYD_1 hairpin. Each value represents the mean ± SD for n=3.

(C) The cellular DHU/uracil ratio increases during EMT. HMLE-Twist-ER cells were treated with OHT for 15 days. At the indicated time points, samples were collected and subjected to LC-MS analysis to determine DHU and uracil levels. The molar concentration ratio between the two metabolites in each sample is presented. Each value represents the mean ± SD for n=3.

(D) The cellular DHU/uracil ratio is elevated in Basal B relative to luminal breast cancer cell lines. The abundance of DHU and uracil was measured in the indicated breast cancer cell lines (green, luminal; red, basal B) using LC-MS. Each value represents the mean ± SD for n=3.

(E) The cellular DHU/uracil ratio is elevated in high-grade relative to low-grade HCC cell lines. Same as (D), but for HCC cell lines (blue, low-grade; brown, high-grade). Each value represents the mean ± SD for n=3.

Figure 6: DPYD activity is essential for the EMT

(A) Mouse DPYD-I560S fails to rescue the effects of DPYD knockdown (KD) on the EMT. HMLE-Twist-ER cells were infected with empty vector, mouse DPYD (mDPYD) or partially active mouse DPYD (DPYD-I560S),

together with either shGFP or shDPYD_1. The cells were treated with OHT for 15 days, as indicated, followed by FACS analysis as in Figure 4A.

(B) Mouse DPYD-I560S fails to rescue the effects of DPYD KD on ZEB1 expression. HMLE-Twist-ER cells infected with the indicated hairpins and vectors were either left untreated or treated with OHT, followed by immunoblotting with the indicated antibodies.

(C) Mouse DPYD-I560S fails to rescue the effects of DPYD KD on mammosphere formation. Cells treated as in (B) were subjected to the in vitro mammosphere formation assay as in Figure 4C.

(D) DPYD activity accelerates the EMT. HMLE-Twist-ER cells expressing shDPYD_1, human DPYD (DPYD-FLAG), mouse DPYD, or partially active human DPYD-FLAG-I560S were either left untreated or treated with OHT for 10 days, followed by FACS analysis as in Figure 4A. The percentage of cells in each gate is presented.

(E) Expression of catalytically attenuated DPYD reduces ZEB1 expression. Cells infected with the indicated constructs were either left untreated or treated with OHT for 10 days, followed by immunoblotting with the indicated antibodies.

(F) DPYD activity enhances mammosphere formation. Cells treated as in (D) were subjected to the in vitro mammosphere formation assay as in Figure 4C.

(G) DPYD products rescue the effect of DPYD KD on mammosphere formation. HMLE-Twist-ER cells expressing shDPYD_1 were treated with the indicated concentrations of dihydrouracil (DHU) or dihydrothymine (DHT) and subjected to the in vitro mammosphere formation assay as in Figure 4C.

Table 1: The mesenchymal metabolic signature genes, classified by metabolic pathway.

Figure S1: High-grade cancer cell lines co-cluster with mesenchymal cells based on metabolic gene expression

(A) The mesenchymal group is composed of cell lines from diverse origins. The pie chart represents the proportion of cell lines of each type in the mesenchymal group (defined by the clustering in Figure 1A). For each cancer type, the number of cell lines falling into the mesenchymal cluster relative to the total number of cell lines of that cancer type in the database is indicated as a fraction.

(B) HCC cell lines can be classified as low or high grade based on the expression levels of liver-specific genes. The heatmap represents liver-specific gene expression in normal liver tissue, HCC cell lines, and primary HCC tumors (separated by gray lines). The arrays were normalized to the normal tissue median of medians and subject to array-based hierarchical clustering. Color bar shows Log_2 scale.

(C) Neither NSCLC or colon cancer cell line subtypes cluster in the mesenchymal group. Each row represents one subtype of NSCLC or colon cancer (similar to Figures 1B and 1C). Within each row, each black line represents the position of a cell line. The order of the cell lines is identical to Figure 1.

(D) Cancer cell lines classified as mesenchymal-like based on metabolic gene expression (Figure 1A) display increased expression of known EMT markers. Gene-set enrichment analysis (GSEA) was applied to all genes, ranked based on the relative expression between the cell lines falling into the mesenchymal cluster (Figure 1A) and all other cell lines in the dataset. The FDR q-value was computed by GSEA.

Figure S2: Expression of MMS genes correlates with that of known mesenchymal markers in cancer cell lines and primary tumors

(A) The MMS genes are co-expressed with EMT markers in primary tumors. Two-way hierarchical clustering of 1,460 primary tumors was performed based on the expression levels of the 44 MMS genes. The values represent the $\log_2$ ratio of each expression value to the primary tumor median of medians (see Materials and Methods). The upper panel represents the MMS gene expression and the lower panel represents the expression of known mesenchymal and epithelial markers in all the primary tumors. Color bar shows Log_2 scale.

(B) Individual validation of known epithelial and mesenchymal marker expression levels by quantitative real-time PCR (qPCR) in breast cancer and HCC cell lines. Each value represents the mean ± SEM for n=3.

(C) Induction of EMT by activation of ectopic Twist expression in HMLE-Twist-ER cells. HMLE-Twist-ER cells were treated with hydroxytamoxifen (OHT) for 15 days. Every three days, samples were collected and subjected to FACS analysis using CD24-FITC and CD44-APC. The percentage of epithelial and mesenchymal cells was determined.

(D) FACS profile of untreated HMLE-Twist-ER cells and Naturally arising mesenchymal cells (NAMEC) cells, which are mesenchymal cells derived from HMLE. Both cells lines were subjected to FACS analysis using CD24-FITC and CD44-APC.

Figure S3: A FACS-based pooled shRNA screen

(A) Genes targeted by shRNAs included in the screen, listed by sub-pool. The MMS list contains only 42 genes because NT5E was considered a known mesenchymal gene, and no hairpins were available for ENPP1.

(B) FACS sorting gates used in the screen. Untreated HMLE-Twist-ER cells (top), NAMEC cells (middle), or HMLE-Twist-ER cells infected by the shRNA library and treated with OHT for 15 days were stained for CD24 and CD44 expression and subjected to FACS.

Figure S4: DPYD expression is essential for the EMT

(A) DPYD expression negatively correlates with the proportion of epithelial cells. HMLE-Twist-ER cells were infected with a variety of hairpins against DPYD and treated with OHT for 15 days. The DPYD expression level was measured by qPCR, and the percentage of cells remaining in the epithelial state was determined by FACS analysis using CD24 and CD44 as markers to separate the epithelial and mesenchymal populations.

(B) DPYD hairpins strongly reduce DPYD expression. HMLE-Twist-ER cells were infected with the indicated hairpins and DPYD expression levels were measured by qPCR.

(C) DPYD knockdown does not affect proliferation. HMLE-Twist-ER cells were infected with the indicated hairpins and the proliferation rate was measured using CellTiterGlo. The number of cells at each time point is represented by relative light units (RLU)(Y-axis), by days (X-axis).

(D) DPYD knockdown reduces ZEB1 expression level. HMLE-Twist-ER cells were infected with the indicated hairpins and left untreated or treated with OHT. The cells were treated with OHT for 15 days and the ZEB1 expression level was measured using qPCR

(E) Sequence alignment between human and mouse *DPYD* in the region of the human gene targeted by shDPYD_1.

Figure S5: DPYD products are elevated in mesenchymal cells

(A) NAMEC cells contain a higher ratio of the DPYD products dihydrothymine (DHT) and dihydrouracil (DHU) to the corresponding substrates, thymine and uracil, as compared to uninduced HMLE-Twist-ER cells. The abundance of all four metabolites was measured by LC-MS in uninduced HMLE-Twist-ER (HMLE, gray) and NAMEC (black) cells. The bars represent the ratio between the two indicated metabolites in each cell line. Each value represents the mean ± SD for n=3. The P values for the indicated comparisons were determined using Student's T test. Each value represents the mean ± SD for n=3. The P values for the indicated comparisons were determined using Student's T test.

(B) DPYD is the only pyrimidine degradation pathway enzyme expressed in HMLE-Twist-ER and NAMEC cell lines. The mRNA from HMLE-Twist-ER cells, NAMEC cells, and human liver was isolated and subjected to qPCR to determine the relative expression of DPYD, DPYS and UPB1. Each value represents the mean ± SEM for n=3.

(C) Expression of DPYD, but not of the other pyrimidine degradation pathway genes, is elevated in Basal B breast and high-grade HCC cancer cell lines. Box plots represent the expression levels of DPYD, DPYS and UPB1 (as indicated) in breast cancer (green, luminal; red, Basal B) and HCC (blue, low-grade; brown, high-grade) subtypes.

Figure S6: DPYD activity is essential for EMT

(A) DHU rescues the effect of DPYD KD on mammosphere formation more strongly than uracil. HMLE-Twist-ER cells expressing shDPYD_1 where either left untreated or induced with OHT with or without the addition of uracil or dihydrouracil, as indicated. The data are reported as the number of mammospheres formed per 500 seeded cells; each value represents the mean ± SD for n=6. The P value for the indicated comparison was determined using Student's T test.

Acknowledgments

We thank all members of the Sabatini Lab for helpful suggestions and Tom DiCesare for graphical assistance. This work was supported by grants from the NIH (CA103866 and AI47389) to D.M.S, the Life Science Research Foundation to Y.D.S., American Cancer Society fellowship to J.R.C, Sally Gordon Fellow of the Damon Runyon cancer research foundation-DRG112-12 to M.E.P., Ruth L Krischstein NRSA F30 fellowship to W.W.C, NIH K99 CA168940 to R.P. and Hope Fund for Cancer Research Fellowship (HFCR-13-03-03) Department of Defense Breast Cancer Research Program (BC123066) to E.F. M.B.Y. is supported by NIH grants CA112967 and ES015339. D.M.S. is an investigator of the Howard Hughes Medical Institute.

Reference

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., Mikulits, W., Brabletz, T., Strand, D., Obrist, P., et al. (2007). The transcription factor ZEB1 (δEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene *26*, 6979–6988.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA *100*, 3983–3988.

Amary, M.F., Bacsi, K., Maggiani, F., Damato, S., Halai, D., Berisha, F., Pollock, R., O'Donnell, P., Grigoriadis, A., Diss, T., et al. (2011). IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. J. Pathol. *224*, 334–343.

Amstutz, U., Froehlich, T.K., and Largiadèr, C.R. (2011). Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity. Pharmacogenomics *12*, 1321–1336.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res *35*, D760–D765.

Brabletz, T. (2012). To differentiate or not — routes towards metastasis. Nat Rev Cancer *12*, 425–436.

Büchel, B., Rhyn, P., Schürch, S., Bühr, C., Amstutz, U., and R Largiadèr, C. (2012). LC-MS/MS method for simultaneous analysis of uracil, 5,6-dihydrouracil, 5-fluorouracil and 5-fluoro-5,6-dihydrouracil in human plasma for therapeutic drug monitoring and toxicity prediction in cancer patients. Cancer Sci *27*, 7–16.

Cantor, J.R., and Sabatini, D.M. (2012). Cancer cell metabolism: one hallmark, many faces. Cancer Discov *2*, 881–898.

Carey, L., Winer, E., Viale, G., Cameron, D., and Gianni, L. (2010). Triple-negative breast cancer: disease entity or title of convenience? Nature Reviews Clinical Oncology *7*, 683–692.

Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D'Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013). Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. Cell *154*, 61–74.

Chaneton, B., Hillmann, P., Zheng, L., Martin, A.C.L., Maddocks, O.D.K., Chokkathukalam, A., Coyle, J.E., Jankevics, A., Holding, F.P., Vousden, K.H., et al. (2012). Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature *491*, 458–462.

Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res *33*, e175.

Dolinnaya, N.G., Kubareva, E.A., Romanova, E.A., Trikin, R.M., and Oretskaya, T.S. (2013). Thymidine glycol: the effect on DNA molecular structure and enzymatic processing. Biochimie *95*, 134–147.

Dontu, G. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev. *17*, 1253–1270.

Elbein, S.C., Zheng, H., Jia, Y., Chu, W., Cooper, J.J., Hale, T., and Zhang, Z. (2004). Molecular screening of the human glutamine–fructose-6-phosphate amidotransferase 1 (GFPT1) gene and association studies with diabetes and diabetic nephropathy. Mol Genet Metab *82*, 321–328.

Elenbaas, B. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. *15*, 50–65.

Ezzeldin, H., and Diasio, R. (2004). Dihydropyrimidine dehydrogenase deficiency, a pharmacogenetic syndrome associated with potentially life-threatening toxicity following 5-fluorouracil administration. Clin Colorectal Cancer *4*, 181–189.

Ferry, G., Moulharat, N., Pradere, J.P., Desos, P., Try, A., Genton, A., Giganti, A., Beucher-Gaudin, M., Lonchampt, M., Bertrand, M., et al. (2008). S32826, A Nanomolar Inhibitor of Autotaxin: Discovery, Synthesis and Applications as a Pharmacological Tool. Journal of Pharmacology and Experimental Therapeutics *327*, 809–819.

Gottlieb, E., and Tomlinson, I.P.M. (2005). Mitochondrial tumour suppressors: a genetic and biochemical update. Nat Rev Cancer *5*, 857–866.

H Ide, S.S.W. (1988). Dihydrothymidine and thymidine glycol triphosphates as substrates for DNA polymerases: differential recognition of thymine C5-C6 bond saturation and sequence specificity of incorporation. Nucleic Acids Res *16*, 11339.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

He, W., Miao, F.J.P., Lin, D.C.H., Schwandner, R.T., Wang, Z., Gao, J., Chen, J.-L., Tian, H., and Ling, L. (2004). Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors. Nat Cell Biol *429*, 188–193.

Hensley, C.T., Wasti, A.T., and Deberardinis, R.J. (2013). Glutamine and cancer: cell biology, physiology, and clinical opportunities. J. Clin. Invest. *123*, 3678–3684.

Hsu, P.P., and Sabatini, D.M. (2008). Cancer cell metabolism: Warburg and beyond. Cell *134*, 703–707.

Hu, J., Locasale, J.W., Bielas, J.H., O'Sullivan, J., Sheahan, K., Cantley, L.C., Heiden, M.G.V., and Vitkup, D. (2013). Heterogeneity of tumor-induced gene expression changes in the human metabolic network. Nat Biotechnol.

Ide, H., Melamede, R.J., and Wallace, S.S. (1987). Synthesis of dihydrothymidine and thymidine glycol 5'-triphosphates and their ability to serve as substrates for Escherichia coli DNA polymerase I. Biochemistry *26*, 964–969.

Jaggupilli, A., and Elkord, E. (2012). Significance of CD44 and CD24 as cancer stem cell markers: an enduring ambiguity. Clin. Dev. Immunol. *2012*, 708036.

Kao, J., Salari, K., Bocanegra, M., Choi, Y.-L., Girard, L., Gandhi, J., Kwei, K.A., Hernandez-Boussard, T., Wang, P., Gazdar, A.F., et al. (2009). Molecular Profiling of Breast Cancer Cell Lines Defines Relevant Tumor Models and Provides a Resource for Cancer Gene Discovery. PLoS ONE *4*, e6146.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. *121*, 2750–2767.

Li, S., Mo, C., Peng, Q., Kang, X., Sun, C., Jiang, K., Huang, L., Lu, Y., Sui, J., Qin, X., et al. (2013). Cell Surface Glycan Alterations in Epithelial Mesenchymal Transition Process of Huh7 Hepatocellular Carcinoma Cell. PLoS ONE *8*, e71273.

Li, Y., VandenBoom, T.G., Kong, D., Wang, Z., Ali, S., Philip, P.A., and Sarkar, F.H. (2009). Up-regulation of miR-200 and let-7 by Natural Agents Leads to the Reversal of Epithelial-to-Mesenchymal Transition in Gemcitabine-Resistant Pancreatic Cancer Cells. Cancer Res *69*, 6704–6712.

Locasale, J.W., Grassian, A.R., Melman, T., Lyssiotis, C.A., Mattaini, K.R., Bass, A.J., Heffron, G., Metallo, C.M., Muranen, T., Sharfi, H., et al. (2011). Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to

oncogenesis. Br J Cancer *43*, 869–874.

Lohkamp, B., Voevodskaya, N., Lindqvist, Y., and Dobritzsch, D. (2010). Insights into the mechanism of dihydropyrimidine dehydrogenase from site-directed mutagenesis targeting the active site loop and redox cofactor coordination. Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics *1804*, 2198–2206.

Ma, Z., and Vosseller, K. (2013). O-GlcNAc in cancer biology. Amino Acids 1–15.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. Cell *133*, 704–715.

Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D., et al. (2009). Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. N Engl J Med *361*, 1058–1066.

Mizutani, Y., Wada, H., Yoshida, O., Fukushima, M., Nakanishi, H., Nakao, M., and Miki, T. (2003). Significance of dihydropyrimidine dehydrogenase activity in renal cell carcinoma. Eur. J. Cancer *39*, 541–547.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 267–273.

Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. PLoS ONE *3*, e2888.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., et al. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515–527.

Nieto, M.A. (2011). The Ins and Outs of the Epithelial to Mesenchymal Transition in Health and Disease. Annu. Rev. Cell Dev. Biol. *27*, 347–376.

Nieto, M.A., and Cano, A. (2012). The epithelial–mesenchymal transition under control: Global programs to regulate epithelial plasticity. Seminars in Cancer Biology *22*, 361–368.

Nomura, D.K., Long, J.Z., Niessen, S., Hoover, H.S., Ng, S.-W., and Cravatt, B.F. (2010). Monoacylglycerol Lipase Regulates a Fatty Acid Network that Promotes Cancer Pathogenesis. Cell *140*, 49–61.

Offer, S.M., Wegner, N.J., Fossum, C., Wang, K., and Diasio, R.B. (2013). Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. Cancer Res.

Onganer, P.U., Seckl, M.J., and Djamgoz, M.B.A. (2005). Neuronal characteristics of small-cell lung cancer. Br J Cancer *93*, 1197–1201.

Parham, D.M. (2001). Neuroectodermal and neuroendocrine tumors principally seen in children. Am. J. Clin. Pathol. *115 Suppl*, S113–S128.

Park, J.G., Lee, J.H., Kang, M.S., Park, K.J., Jeon, Y.M., Lee, H.J., Kwon, H.S., Park, H.S., Yeo, K.S., and Lee, K.U. (1995). Characterization of cell lines established from human hepatocellular carcinoma. Int J Cancer *62*, 276–282.

Park, S.Y., Kim, H.S., Kim, N.H., Ji, S., Cha, S.Y., Kang, J.G., Ota, I., Shimada, K., Konishi, N., Nam, H.W., et al. (2010). Snail1 is stabilized by O-GlcNAc modification in hyperglycaemic condition. Embo J *29*, 3787–3796.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G.L., et al. (2008). An Integrated Genomic Analysis of Human Glioblastoma Multiforme. Science *321*, 1807–1812.

Peinado, H.E.C., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer *7*, 415–428.

Possemato, R., Marks, K.M., Shaul, Y.D., Pacold, M.E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.-K., Jang, H.G., Jha, A.K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature.

Samadi, N., Bekele, R., Capatos, D., Venkatraman, G., Sariahmetoglu, M., and Brindley, D.N. (2011). Regulation of lysophosphatidate signaling by autotaxin and lipid phosphate phosphatases with respect to tumor progression, angiogenesis, metastasis and chemo-resistance. Biochimie *93*, 61–70.

Scheel, C., and Weinberg, R.A. (2011). Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? Int J Cancer *129*, 2310–2314.

Schulze, A., and Harris, A.L. (2012). How cancer metabolism is tuned for proliferation and vulnerable to disruption. Nature *491*, 364–373.

Singh, A., and Settleman, J. (2010). EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. Oncogene *29*, 4741–4751.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette,

M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Application of a priori established gene sets to discover biologically important differential expression in microarray data. Proceedings of the National Academy of Sciences *102*, 15278–15279.

Tam, W.L., Lu, H., Buikhuisen, J., Soh, B.S., Lim, E., Reinhardt, F., Wu, Z.J., Krall, J.A., Bierie, B., Guo, W., et al. (2013). Protein Kinase C α Is a Central Signaling Node and Therapeutic Target for Breast Cancer Stem Cells. Cancer Cell *24*, 347–364.

Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., Hartwell, K., Onder, T.T., Gupta, P.B., Evans, K.W., et al. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci USA *107*, 15449–15454.

Tennant, D.A., n, R.U.L.V.D.A., and Gottlieb, E. (2010). Targeting metabolic transformation for cancer therapy. Nat Rev Cancer *10*, 267–277.

Ulanovskaya, O.A., Zuhl, A.M., and Cravatt, B.F. (2013). NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. Nature Chemical Biology.

Van Gennip, A.H., and Van Kuilenburg, A.B. (2000). Defects of Pyrimidine Degradation: Clinical, Molecular and Diagnostic Aspects. Adv. Exp. Med. Biol. *486*, 233–241.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110.

Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. Cancer Cell *21*, 297–308.

Yoo, B.K., Gredler, R., Vozhilla, N., Su, Z.Z., Chen, D., Forcier, T., Shah, K., Saxena, U., Hansen, U., Fisher, P.B., et al. (2009). Identification of genes conferring resistance to 5-fluorouracil. Proceedings of the National Academy of Sciences *106*, 12938–12943.

Zhang, H. (2004). Common Variants in Glutamine:Fructose-6-Phosphate Amidotransferase 2 (GFPT2) Gene Are Associated with Type 2 Diabetes, Diabetic Nephropathy, and Increased GFPT2 mRNA Levels. Journal of Clinical Endocrinology & Metabolism *89*, 748–755.

Zhang, W.C., Shyh-Chang, N., Yang, H., Rai, A., Umashankar, S., Ma, S., Soh, B.S., Sun, L.L., Tai, B.C., Nga, M.E., et al. (2012). Glycine Decarboxylase Activity Drives Non-Small Cell Lung Cancer Tumor-Initiating Cells and Tumorigenesis. Cell *148*, 259–272.

Zhi, X., Wang, Y., Yu, J., Yu, J., Zhang, L., Yin, L., and Zhou, P. (2012). Potential prognostic biomarker CD73 regulates epidermal growth factor receptor expression in human breast cancer. IUBMB Life *64*, 911–920.

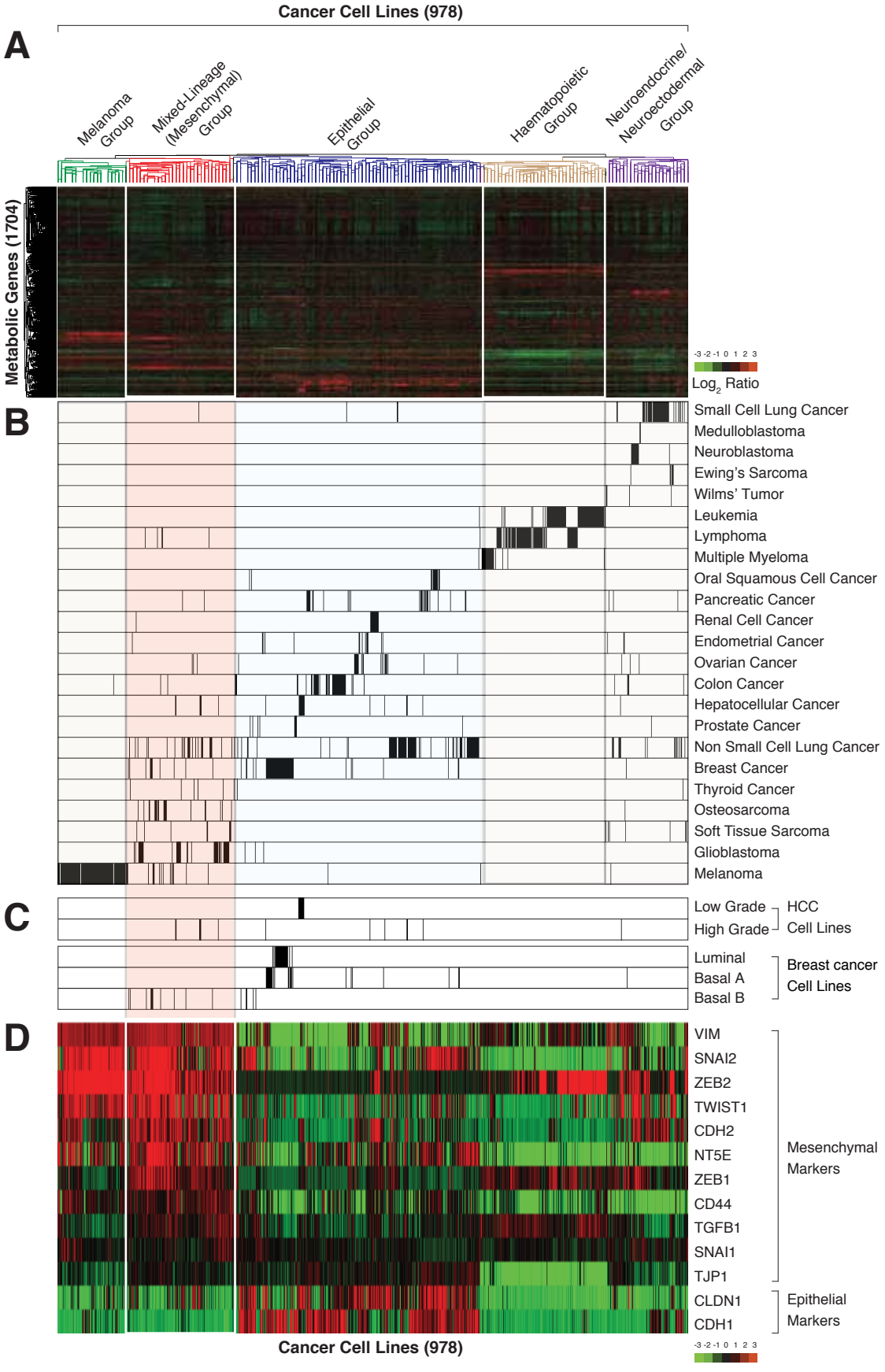

Cancer Cell Lines (978)
A
Melanoma Group
Mixed-Lineage (Mesenchymal) Group
Epithelial Group
Haematopoietic Group
Neuroendocrine/ Neuroectodermal Group
Metabolic Genes (1704)
-3 -2 -1 0 1 2 3
Log2 Ratio
B
Small Cell Lung Cancer
Medulloblastoma
Neuroblastoma
Ewing's Sarcoma
Wilms' Tumor
Leukemia
Lymphoma
Multiple Myeloma
Oral Squamous Cell Cancer
Pancreatic Cancer
Renal Cell Cancer
Endometrial Cancer
Ovarian Cancer
Colon Cancer
Hepatocellular Cancer
Prostate Cancer
Non Small Cell Lung Cancer
Breast Cancer
Thyroid Cancer
Osteosarcoma
Soft Tissue Sarcoma
Glioblastoma
Melanoma
C
Low Grade
High Grade
HCC Cell Lines
Luminal
Basal A
Basal B
Breast cancer Cell Lines
D
VIM
SNAI2
ZEB2
TWIST1
CDH2
NT5E
ZEB1
CD44
TGFB1
SNAI1
TJP1
Mesenchymal Markers
CLDN1
CDH1
Epithelial Markers
Cancer Cell Lines (978)
-3 -2 -1 0 1 2 3

Figure 1

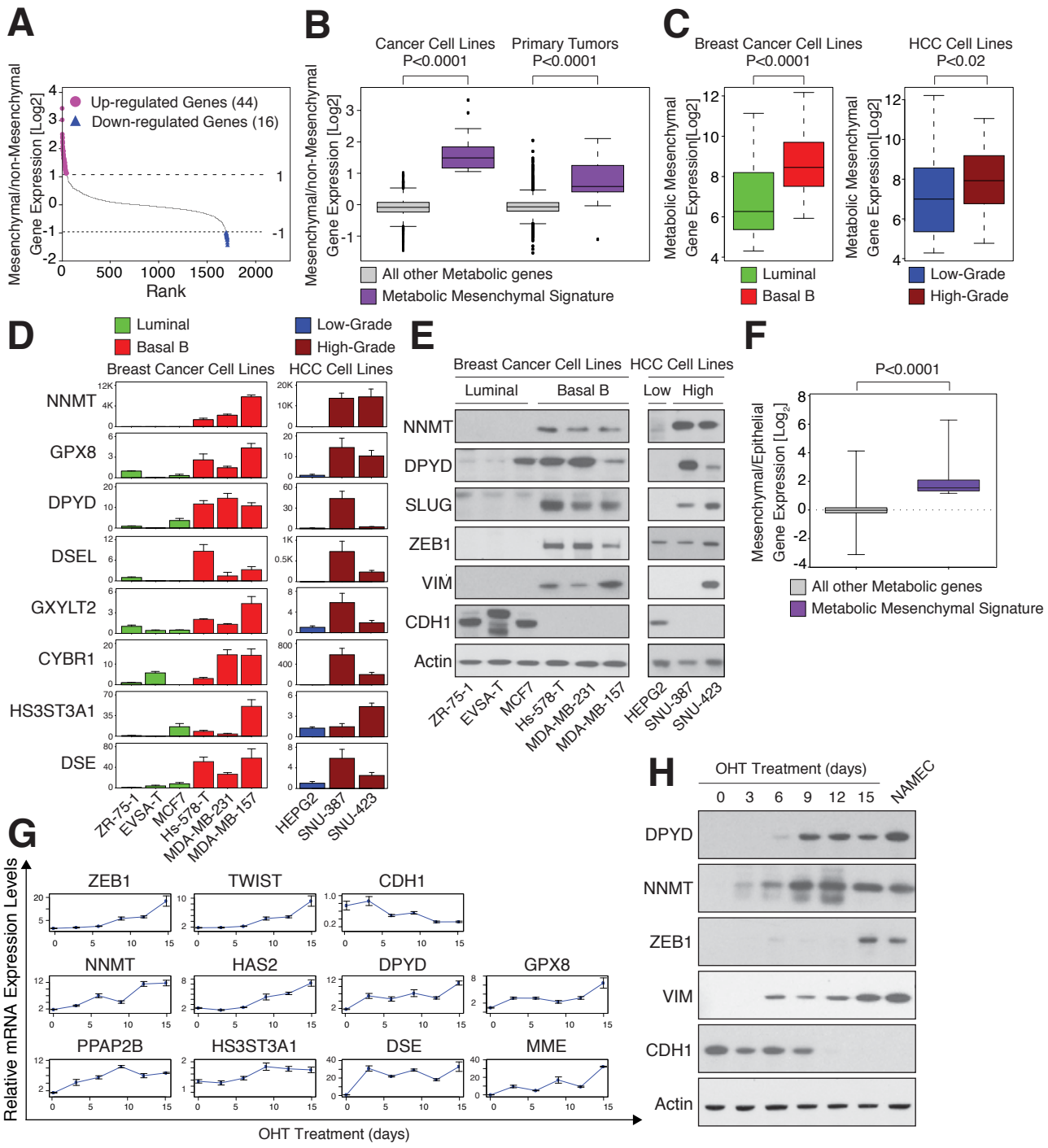
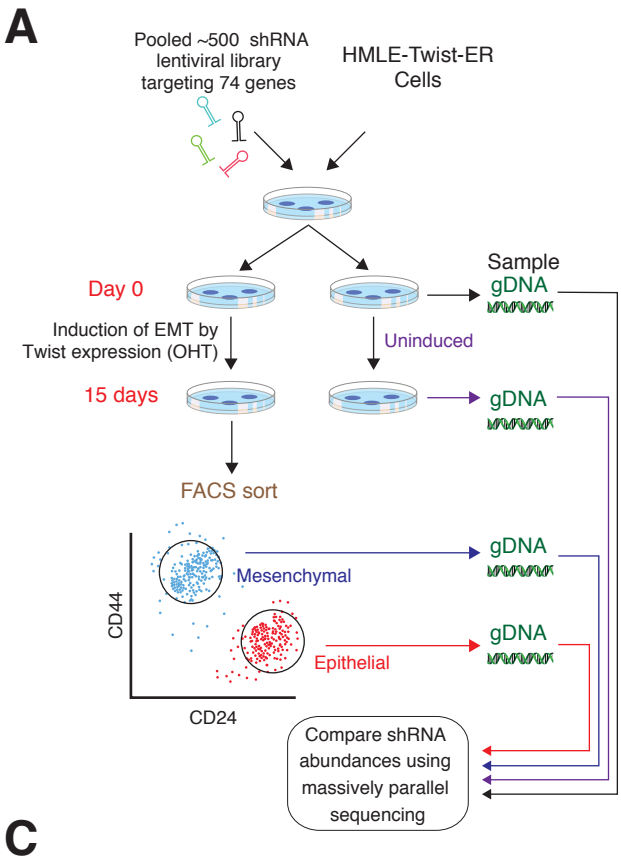
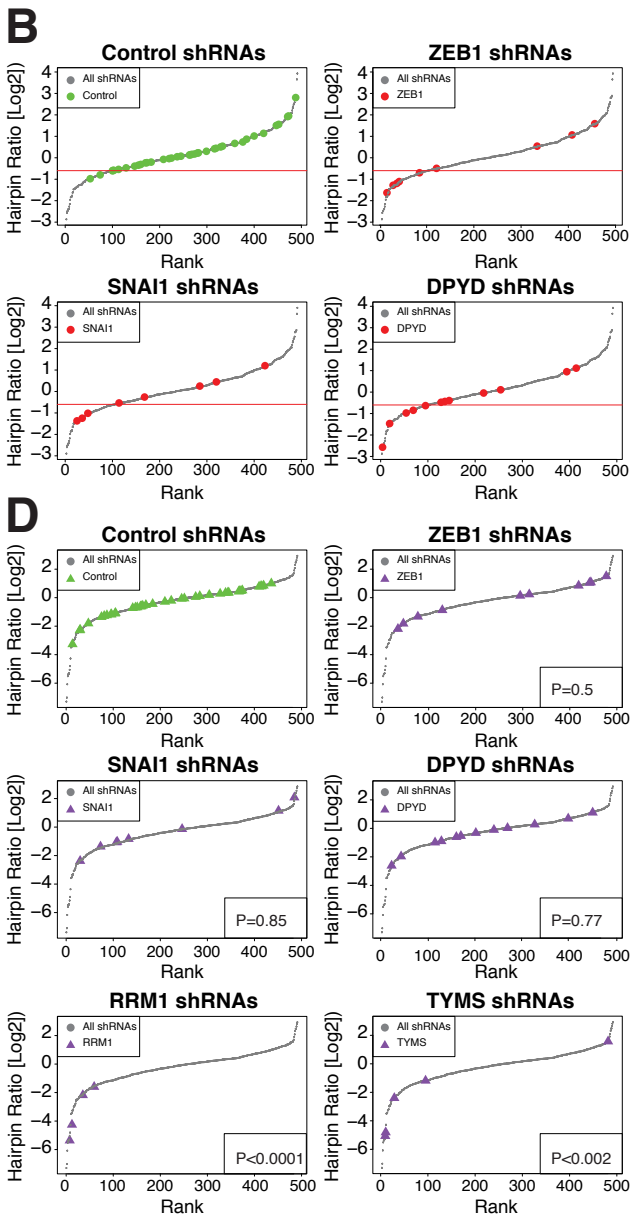

A
Mesenchymal/non-Mesenchymal Gene Expression [Log2]
Up-regulated Genes (44)
Down-regulated Genes (16)
Rank
B
Cancer Cell Lines P<0.0001
Primary Tumors P<0.0001
All other Metabolic genes
Metabolic Mesenchymal Signature
C
Breast Cancer Cell Lines P<0.0001
HCC Cell Lines P<0.02
Metabolic Mesenchymal Gene Expression[Log2]
Luminal
Basal B
Low-Grade
High-Grade
D
Luminal
Basal B
Low-Grade
High-Grade
Breast Cancer Cell Lines
HCC Cell Lines
NNMT
GPX8
DPYD
DSEL
GXYLT2
CYBR1
HS3ST3A1
DSE
ZR-75-1
EVSA-T
MCF7
Hs-578-T
MDA-MB-231
MDA-MB-157
HEPG2
SNU-387
SNU-423
E
Breast Cancer Cell Lines
HCC Cell Lines
Luminal
Basal B
Low
High
NNMT
DPYD
SLUG
ZEB1
VIM
CDH1
Actin
F
P<0.0001
Mesenchymal/Epithelial Gene Expression [Log2]
All other Metabolic genes
Metabolic Mesenchymal Signature
G
Relative mRNA Expression Levels
ZEB1
TWIST
CDH1
NNMT
HAS2
DPYD
GPX8
PPAP2B
HS3ST3A1
DSE
MME
OHT Treatment (days)
H
OHT Treatment (days)
0 3 6 9 12 15 NAMEC
DPYD
NNMT
ZEB1
VIM
CDH1
Actin

Figure 2





C

Sub-Pools	Number of Hit Genes /Total Genes
MMS genes	16/42
Non-MMS Genes	2/16
Known Mesenchymal Genes	6/12

MMS Hit Genes

DPYD	GPX8	EXT1	HS3ST3A1
PPAP2B	MGST1	DSEL	MSRB3
MICAL2	PLCB4	B3GNT9	SULF1
PPAPDC1A	COX7A1	CA12	UAP1

Figure 3

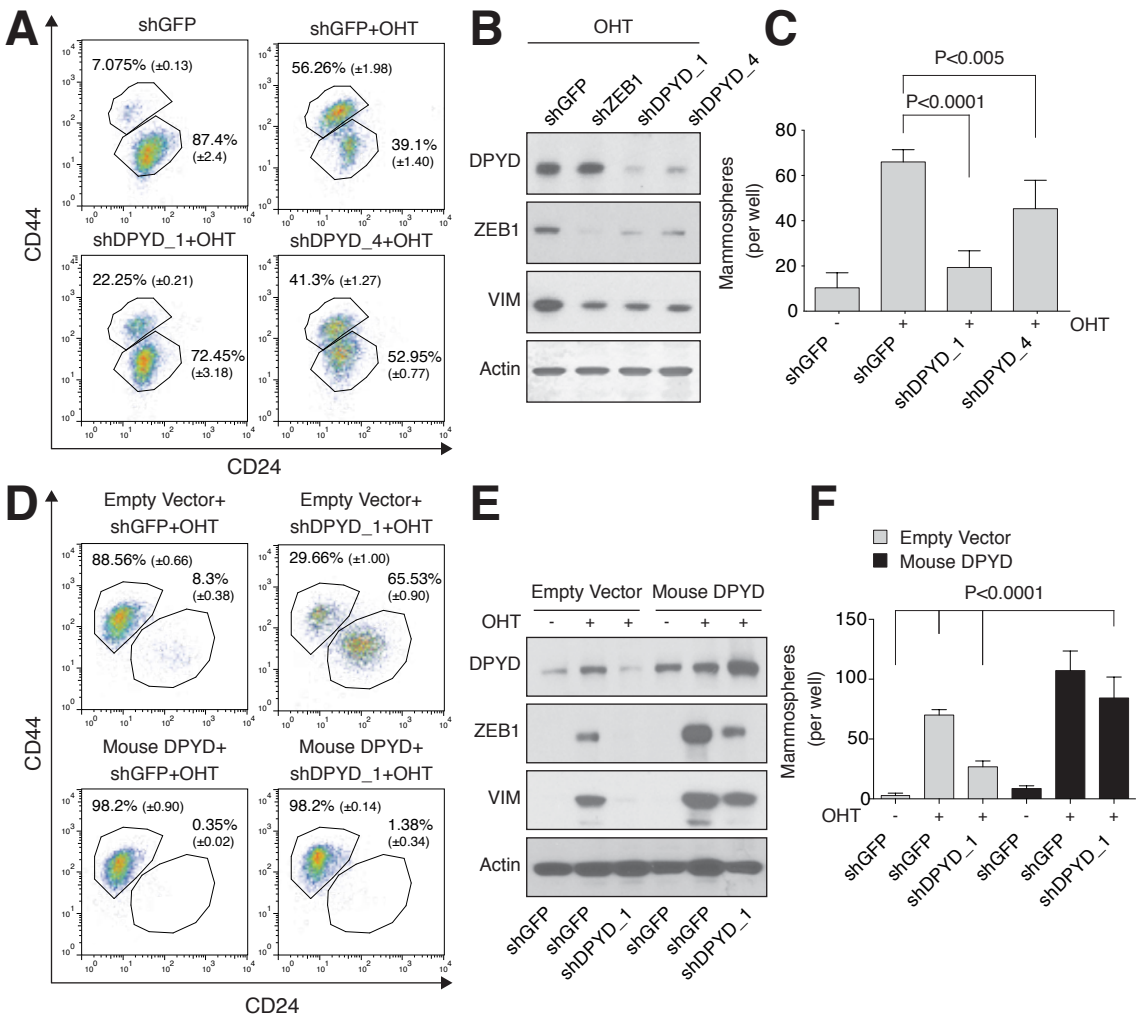

A
shGFP
shGFP+OHT
7.075% (±0.13)
87.4% (±2.4)
56.26% (±1.98)
39.1% (±1.40)
shDPYD_1+OHT
shDPYD_4+OHT
22.25% (±0.21)
72.45% (±3.18)
41.3% (±1.27)
52.95% (±0.77)
CD44
CD24
B
OHT
shGFP
shZEB1
shDPYD_1
shDPYD_4
DPYD
ZEB1
VIM
Actin
C
P<0.005
P<0.0001
Mammospheres (per well)
OHT
shGFP
shGFP
shDPYD_1
shDPYD_4
D
Empty Vector+ shGFP+OHT
Empty Vector+ shDPYD_1+OHT
88.56% (±0.66)
8.3% (±0.38)
29.66% (±1.00)
65.53% (±0.90)
Mouse DPYD+ shGFP+OHT
Mouse DPYD+ shDPYD_1+OHT
98.2% (±0.90)
0.35% (±0.02)
98.2% (±0.14)
1.38% (±0.34)
CD44
CD24
E
Empty Vector
Mouse DPYD
OHT
DPYD
ZEB1
VIM
Actin
shGFP
shGFP
shDPYD_1
F
Empty Vector
Mouse DPYD
P<0.0001
Mammospheres (per well)
OHT
shGFP
shGFP
shDPYD_1
shGFP
shGFP
shDPYD_1

Figure 4

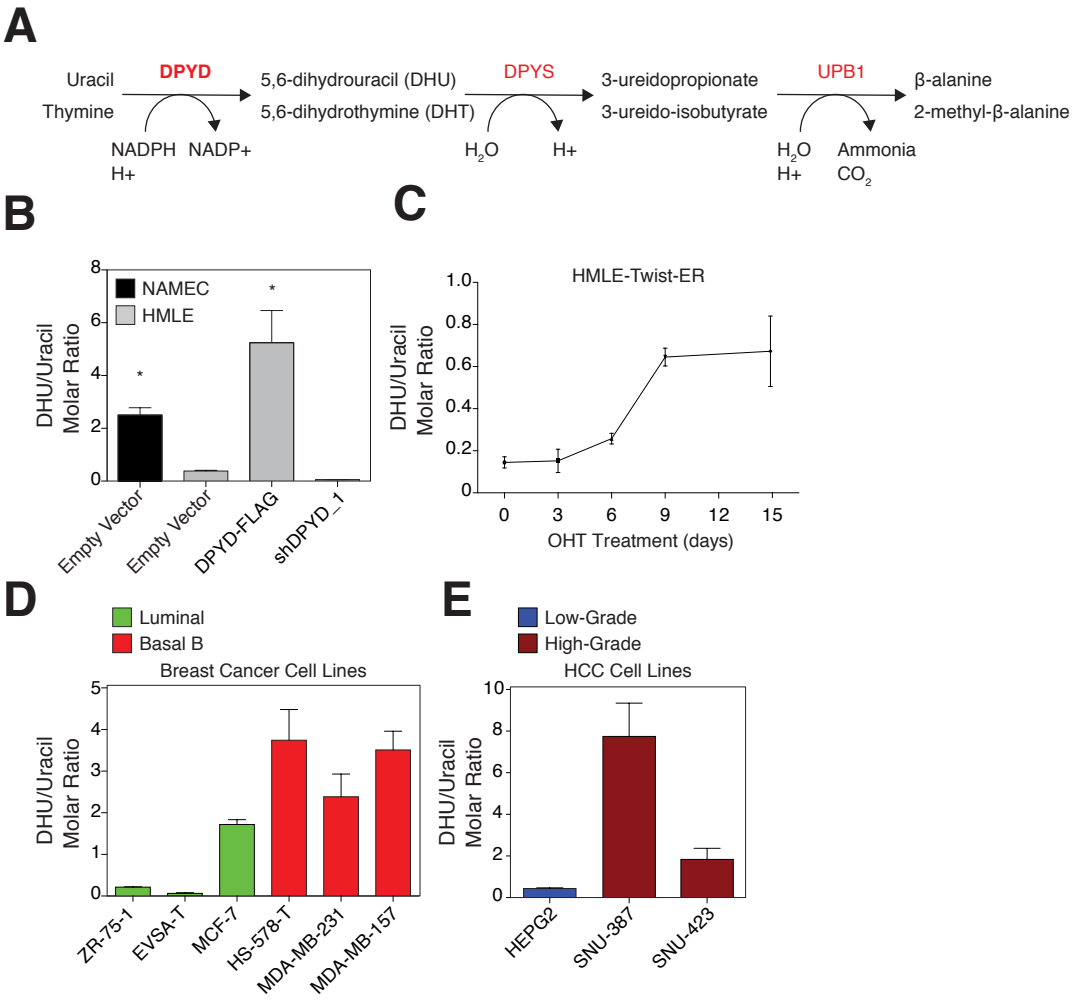

A
Uracil
Thymine
DPYD
NADPH
H+
NADP+
5,6-dihydrouracil (DHU)
5,6-dihydrothymine (DHT)
DPYS
H2O
H+
3-ureidopropionate
3-ureido-isobutyrate
UPB1
H2O
H+
Ammonia
CO2
β-alanine
2-methyl-β-alanine
B
NAMEC
HMLE
DHU/Uracil Molar Ratio
Empty Vector
Empty Vector
DPYD-FLAG
shDPYD_1
C
HMLE-Twist-ER
DHU/Uracil Molar Ratio
OHT Treatment (days)
D
Luminal
Basal B
Breast Cancer Cell Lines
DHU/Uracil Molar Ratio
ZR-75-1
EVSA-T
MCF-7
HS-578-T
MDA-MB-231
MDA-MB-157
E
Low-Grade
High-Grade
HCC Cell Lines
DHU/Uracil Molar Ratio
HEPG2
SNU-387
SNU-423

Figure 5

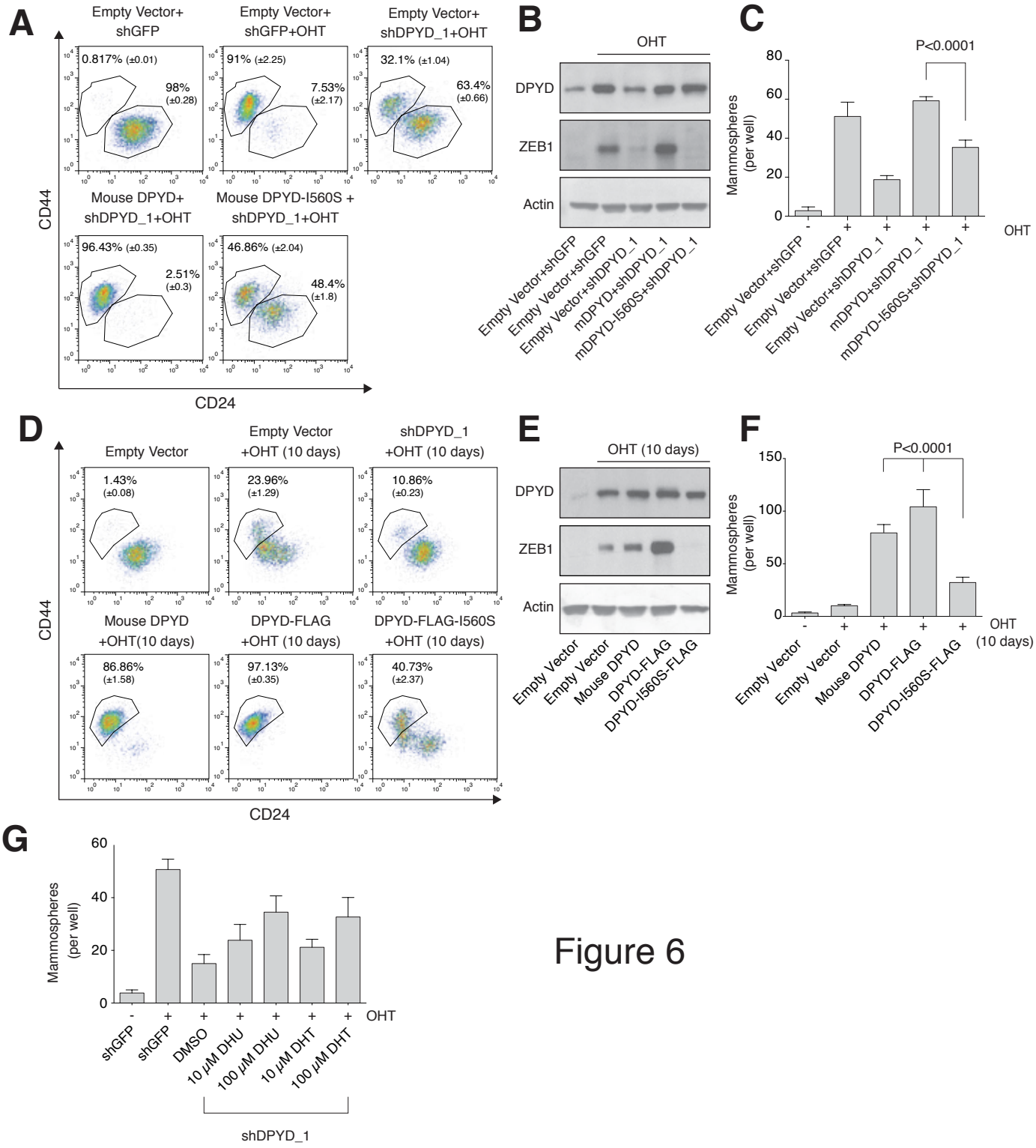
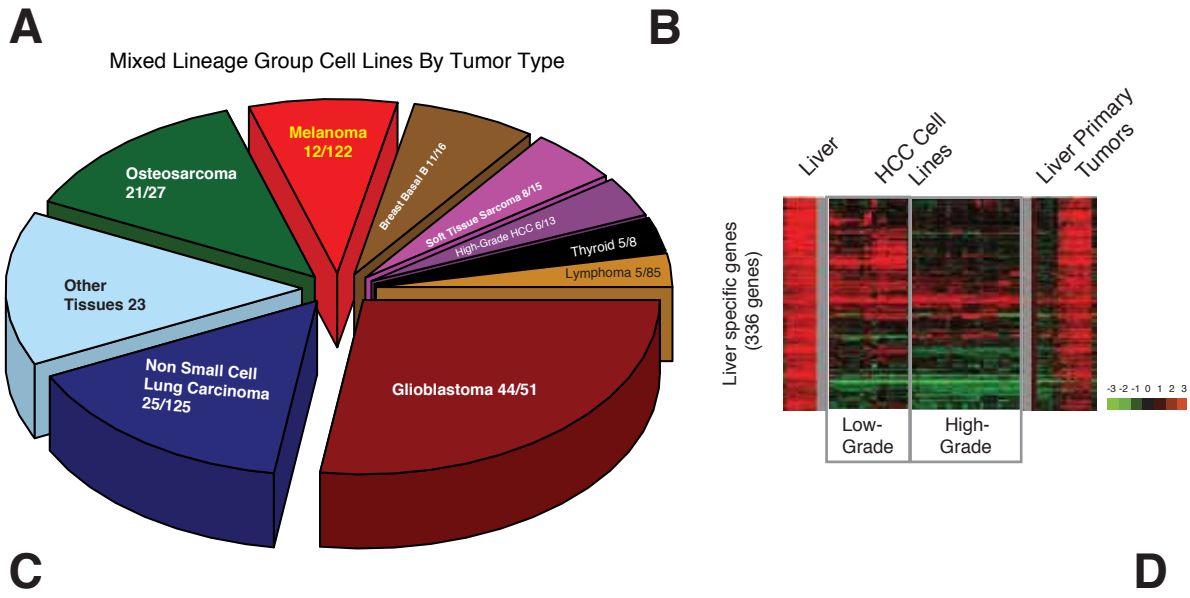

A
Empty Vector+ shGFP
Empty Vector+ shGFP+OHT
Empty Vector+ shDPYD_1+OHT
0.817% (±0.01)
98% (±0.28)
91% (±2.25)
7.53% (±2.17)
32.1% (±1.04)
63.4% (±0.66)
Mouse DPYD+ shDPYD_1+OHT
Mouse DPYD-I560S + shDPYD_1+OHT
96.43% (±0.35)
2.51% (±0.3)
46.86% (±2.04)
48.4% (±1.8)
CD44
CD24
B
OHT
DPYD
ZEB1
Actin
Empty Vector+shGFP
Empty Vector+shGFP
Empty Vector+shDPYD_1
mDPYD+shDPYD_1
mDPYD-I560S+shDPYD_1
C
P<0.0001
Mammospheres (per well)
OHT
D
Empty Vector
Empty Vector +OHT (10 days)
shDPYD_1 +OHT (10 days)
1.43% (±0.08)
23.96% (±1.29)
10.86% (±0.23)
Mouse DPYD +OHT(10 days)
DPYD-FLAG +OHT (10 days)
DPYD-FLAG-I560S +OHT (10 days)
86.86% (±1.58)
97.13% (±0.35)
40.73% (±2.37)
CD44
CD24
E
OHT (10 days)
DPYD
ZEB1
Actin
Empty Vector
Empty Vector
Mouse DPYD
DPYD-FLAG
DPYD-I560S-FLAG
F
P<0.0001
Mammospheres (per well)
OHT (10 days)
G
Mammospheres (per well)
shGFP
shGFP
DMSO
10 μM DHU
100 μM DHU
10 μM DHT
100 μM DHT
OHT
shDPYD_1

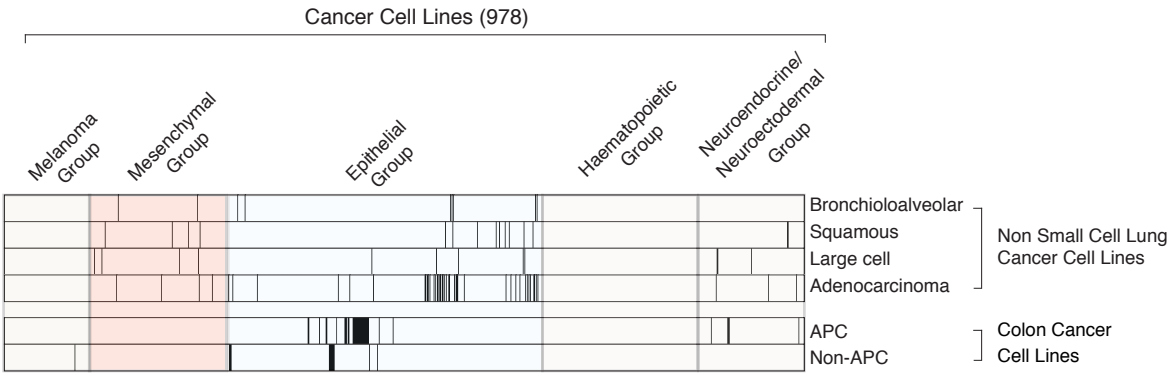
Figure 6

	Metabolic Pathways	Metabolic Enzymes
Nucleotide	Pyrimidine Degradation	DPYD
	Other	AK5, ENPP1, NT5E
Lipids	Triacylglycerol Degradation	MGLL
	Cholesterol	TM7SF2, AKR1B1
	Sphingosine	SPHK1, UGCG
	Signaling	ENPP2, PPAP2B, PPAPDC1A, PDE1C, PLCB4
		PTGR1, PIK3C2B, PLCG2, ALDH1A1, PIP5K1B
Amino Acid	Branched Amino Acid Degradation	BCAT1
	Amino Acid Degradation	CYP1B1
	Tetrahydrobiopterin Biosynthesis	GCH1
Carbon	TCA Cycle	CYBRD, COX7A1, CYBA
Redox	Glutathione	MGST1, GPX8, GPX2
Glycan	Beta-Galactose	ST6GAL1, GLB1L2
	Dermatan Sulfate Biosynthesis	DSE, DSEL
	Heparin Sulfate	HS3ST3A1, EXT1
	Sulfatases	ARSJ, SULF1, PAPSS2
	GlcNAc	GFPT2, GALNT10, UAP1
	Glycan	GXYLT2, GBE1, GLT8D2, GALNT3
	Hyaluronan Synthase	HAS2
	Glyco Protein	PAM, CHI3L1
	Other	B3GNT9, MFNG, HPDL
CoFactor	Nicotinamide	AOX1, NNMT, QPRT
Other		MICAL2, MME, DDAH1, MSRB3, PTER
		CA12, CA2

Underline=Rate Limiting Enzymes Red= MMS Up-regulated Green=MMS Down-regulated

Table 1


A
Mixed Lineage Group Cell Lines By Tumor Type
Melanoma 12/122
Osteosarcoma 21/27
Breast Basal B 11/16
Soft Tissue Sarcoma 8/15
High-Grade HCC 6/13
Thyroid 5/8
Lymphoma 5/85
Other Tissues 23
Non Small Cell Lung Carcinoma 25/125
Glioblastoma 44/51
B
Liver
HCC Cell Lines
Liver Primary Tumors
Liver specific genes (336 genes)
-3 -2 -1 0 1 2 3
Low-Grade
High-Grade


C
Cancer Cell Lines (978)
Melanoma Group
Mesenchymal Group
Epithelial Group
Haematopoietic Group
Neuroendocrine/ Neuroectodermal Group
Bronchioloalveolar
Squamous
Large cell
Adenocarcinoma
Non Small Cell Lung Cancer Cell Lines
APC
Non-APC
Colon Cancer Cell Lines

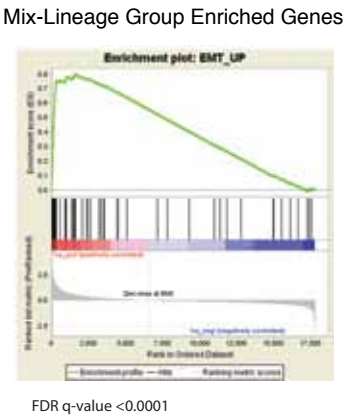

D
Mix-Lineage Group Enriched Genes
Enrichment plot: EMT_UP
FDR q-value <0.0001

Figure S1

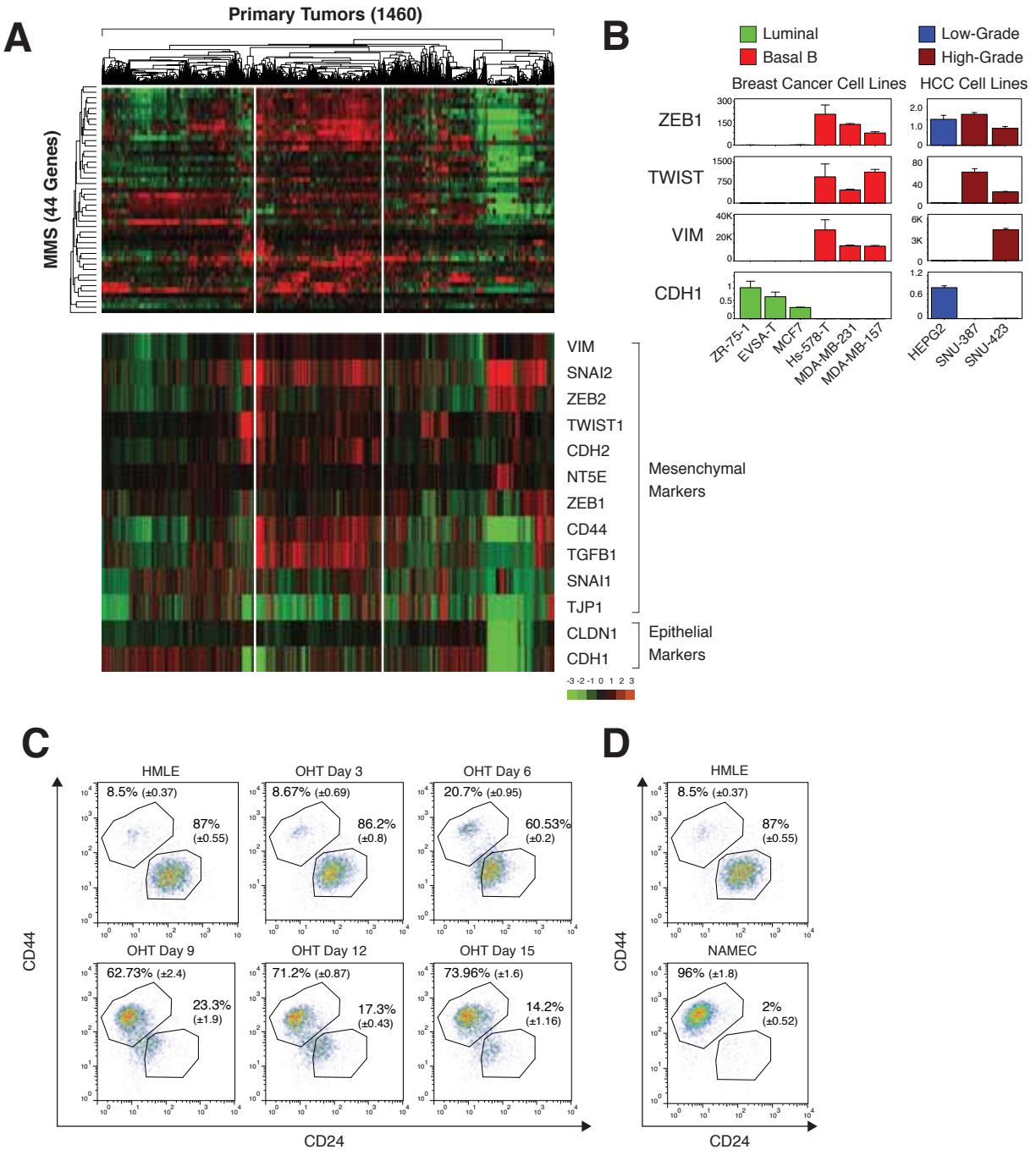

A
Primary Tumors (1460)
MMS (44 Genes)
VIM
SNAI2
ZEB2
TWIST1
CDH2
NT5E
ZEB1
CD44
TGFB1
SNAI1
TJP1
CLDN1
CDH1
Mesenchymal Markers
Epithelial Markers
-3 -2 -1 0 1 2 3
B
Luminal
Basal B
Low-Grade
High-Grade
Breast Cancer Cell Lines
HCC Cell Lines
ZEB1
TWIST
VIM
CDH1
ZR-75-1
EVSA-T
MCF7
Hs-578-T
MDA-MB-231
MDA-MB-157
HEPG2
SNU-387
SNU-423
C
HMLE
OHT Day 3
OHT Day 6
8.5% (±0.37)
87% (±0.55)
8.67% (±0.69)
86.2% (±0.8)
20.7% (±0.95)
60.53% (±0.2)
OHT Day 9
OHT Day 12
OHT Day 15
62.73% (±2.4)
23.3% (±1.9)
71.2% (±0.87)
17.3% (±0.43)
73.96% (±1.6)
14.2% (±1.16)
CD44
CD24
D
HMLE
8.5% (±0.37)
87% (±0.55)
NAMEC
96% (±1.8)
2% (±0.52)
CD44
CD24

Figure S2

A

Control genes	MMS genes		Non-MMS Metabolic Genes	Known Mesenchymal Genes
GFP	AK5	GPX8	ACLY	CD44
LacZ	AKR1B1	GXYLT2	AGK	CDH1
LUCIFERASE	AOX1	HAS2	ALDH1L1	FOXC2
RFP	ARSJ	HS3ST3A1	ARG2	GCG
	B3GNT9	MGLL	CDA	NT5E
	BCAT1	MGST1	DIO1	SNAI1
	CA12	MICAL2	ELOVL5	SNAI2
	CHI3L1	MME	GCNT3	TGFB1
	COX7A1	MSRB3	GGH	TWIST1
	CYBRD1	NNMT	GPX3	VIM
	CYP1B1	PAM	HNMT	ZEB1
	DDAH1	PAPSS2	MAOB	ZEB2
	DPYD	PDE1C	NUDT5	
	DSE	PLCB4	ST6GALNAC2	
	DSEL	PPAP2B	RRM1*	
	ENPP2	PPAPDC1A	TYMS*	
	EXT1	PTGR1		
	GALNT10	SPHK1		
	GBE1	SULF1	* Genes required for proliferation	
	GFPT2	UAP1		
	GLT8D2	UGCG		

B

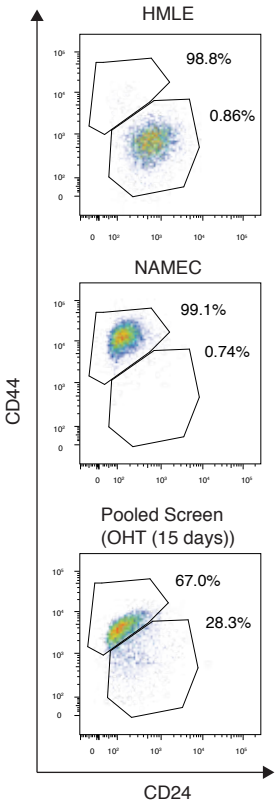


Figure S3

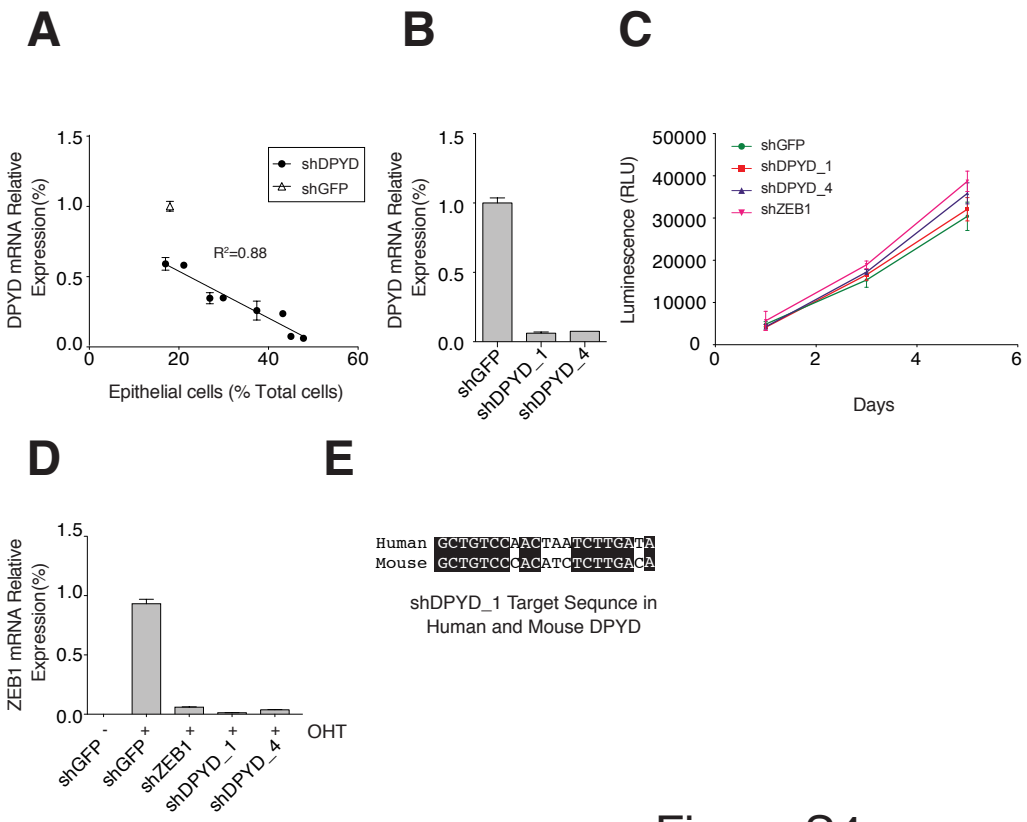

A
B
C
DPYD mRNA Relative Expression(%)
shDPYD
shGFP
R²=0.88
Epithelial cells (% Total cells)
shGFP
shDPYD_1
shDPYD_4
Luminescence (RLU)
shZEB1
Days
D
E
ZEB1 mRNA Relative Expression(%)
OHT
shZEB1
Human GCTGTCCAACTAATCTTGATA
Mouse GCTGTCCGACATCTCTTGACA
shDPYD_1 Target Sequnce in
Human and Mouse DPYD

Figure S4

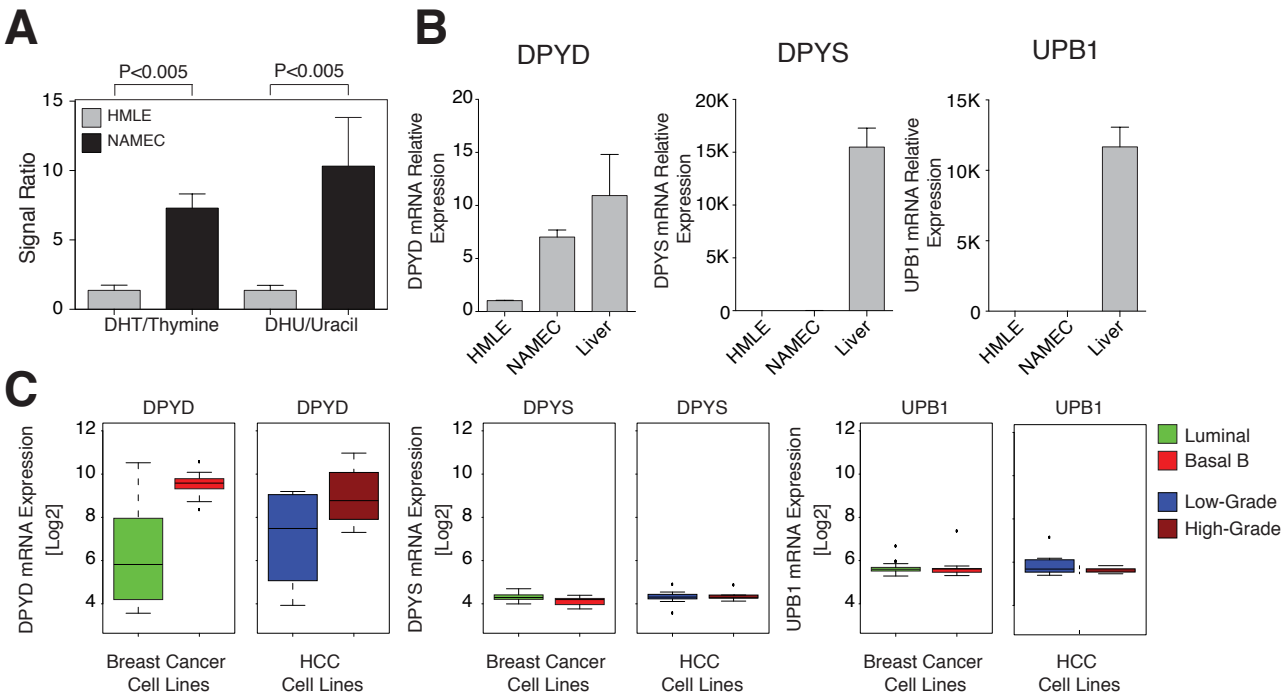

A
P<0.005
P<0.005
HMLE
NAMEC
Signal Ratio
DHT/Thymine
DHU/Uracil
B
DPYD
DPYS
UPB1
DPYD mRNA Relative Expression
DPYS mRNA Relative Expression
UPB1 mRNA Relative Expression
HMLE
NAMEC
Liver
C
DPYD mRNA Expression [Log2]
DPYS mRNA Expression [Log2]
UPB1 mRNA Expression [Log2]
Breast Cancer Cell Lines
HCC Cell Lines
Luminal
Basal B
Low-Grade
High-Grade

Figure S5

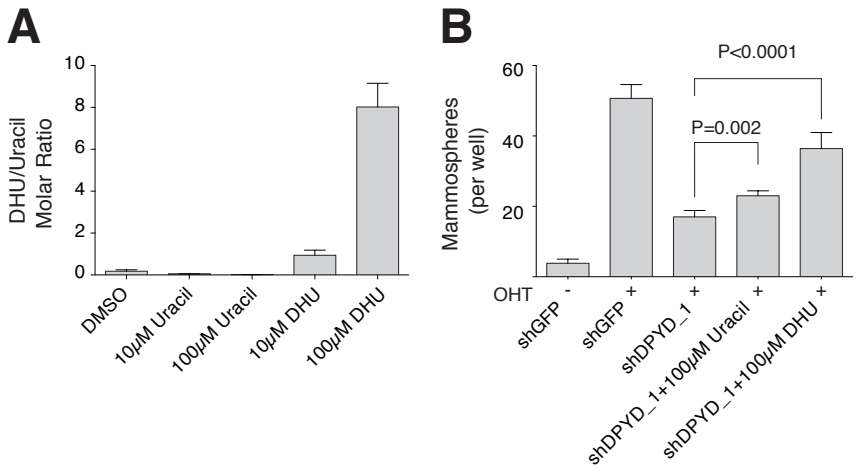

A
DHU/Uracil
Molar Ratio
10
8
6
4
2
0
DMSO
10μM Uracil
100μM Uracil
10μM DHU
100μM DHU
B
Mammospheres
(per well)
60
40
20
P<0.0001
P=0.002
OHT - + + + +
shGFP
shGFP
shDPYD_1
shDPYD_1+100μM Uracil
shDPYD_1+100μM DHU

Figure S6